UNIVERSITY OF CALIFORNIA SAN DIEGO

Nonlinear Modal Methods for Analysis and Synthesis

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Music

by

Joseph Mariglio

Committee in charge:

Professor Miller Puckette, Chair
Professor Henry Abarbanel
Professor Charles Curtis
Professor Mauricio De Oliveira
Professor Tamara Smyth

2018

The dissertation of Joseph Mariglio is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California San Diego

2018

DEDICATION

To spiders.

EPIGRAPH

"Approach what you find repulsive... Go to the places that scare you."

-Machig Labdrön (1055-1149)

TABLE OF CONTENTS

LIST OF SYMBOLS

$A(z^{-1})$ ..... Model output parameter polynomial

$B(z^{-1})$ ..... Model input parameter polynomial (numerator)

$\boldsymbol{C}$ ..... Model process noise parameter

$\gamma(z^{-1})$ ..... Model noise parameter polynomial (numerator)

$D(z^{-1})$ ..... Model noise parameter polynomial (denominator)

$\boldsymbol{E}$ ..Similarity transform from diagonal eigenbasis to real-valued block diagonal basis.

$E(x)$ ..... Expected value

$\boldsymbol{e}_k$ ..... General model residual vector

$\boldsymbol{e}_{\text{exp}}$ ..... The residual vector resulting from the exponential resynthesis step

$\boldsymbol{e}_{\Delta}$ ... The improvement, in decibels, of some iteration of the residual calculation over another

$\boldsymbol{e}_{\text{lin}}$ ..... The residual vector resulting from the linearization step

$F(z^{-1})$ ..... Model input parameter polynomial (denominator)

$\boldsymbol{F}$ ..... N^3-dimensional column vector of quartic cross-terms

$\boldsymbol{F'}$ .....$2\left(\binom{N}{2}+\binom{N}{3}\right)$ dimensional column vector of reduced quartic cross-terms

$\boldsymbol{G}$ .....N^3 x N^3 matrix of quartic terms

$\mathcal{G}$ .....Block of quartic terms

$\boldsymbol{G'}$ ..... $2\left(\binom{N}{2}+\binom{N}{3}\right)x2\left(\binom{N}{2}+\binom{N}{3}\right)$ matrix of reduced quartic terms

$H(z)$ .....Transfer function

$\boldsymbol{H}$ .....Hermitian matrix

K .....Total samples

k ..... Sample index

$\boldsymbol{\Lambda}$ ..... Diagonal, complex, eigenvalue matrix

M .....Number of rows

m ..... Row index

$\boldsymbol{M}$ ..... Diagonal 2×2 block in $\mathcal{M}_{ab}$

μ ..... 1st-order nonlinear model parameter

$\mathcal{M}'$ ..... $2\left(\binom{N}{2}+\binom{N}{3}\right)$ dimensional nonlinear model parameter vector

$N\theta$ ..... Total number of model parameters

N ..... Number of columns

n ..... Column index

$\boldsymbol{N}$ ..... Off-diagonal 2×2 block in $\mathcal{M}_{ab}$

Na ..... Degree of model output parameter polynomial

Nb ..... Degree of model input parameter polynomial (numerator)

Nd ..... Degree of model noise parameter polynomial (denominator)

Nd ..... Degree of model noise parameter polynomial (denominator)

$N\gamma$ ..... Degree of model noise parameter polynomial (numerator)

Φ ..... The index of nonlinearity, in dB

$\boldsymbol{P}_{m,n}$ ..... General model term matrix

$\boldsymbol{Q}_{m,n}$ ..... Quadratic transformation matrix

$\mathcal{R}$ ..... Real-valued $N \times N$ block-rotation matrix

$\boldsymbol{R}$ ..... Real-valued 2×2 rotation matrix $\begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$

$\boldsymbol{S}$ ..... Complex eigenvector basis

$\overset{AB}{\boldsymbol{C}}$ ..... Similarity transforms (above), applied to a tensor (below).

$\boldsymbol{\Sigma}_m$ ..... Measurement noise covariance matrix

$\boldsymbol{\Sigma}_{pm}$ ..... Measurement noise covariance matrix

$\boldsymbol{\Sigma}_p$ ..... Process noise covariance matrix

$\boldsymbol{T}$ ..... State transition matrix

θ_m ..... General model parameter vector

$\boldsymbol{U}$ ..... Similar to a unitary matrix

$u(k)$ ..... Model input signal

$\boldsymbol{V}$ ..... Diagonal basis for antisymmetrizing high-order $\overset{\boldsymbol{SE}}{\mathcal{M}}$

$\boldsymbol{W}$ ..... State-Measurement transformation matrix.

$\xi(k)$ ..... Model noise signal

$x(k)$ ..... Model state variable signal

$\xi m(k)$ ..... Model noise signal

$\xi p(k)$ ..... Model noise signal

$y(k)$ ..... Model output signal

$\mathbb{Z}$ ..... Set of all Integers

z ..... Complex variable

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Miller Puckette for his guidance. Despite life's many attempts to pull me off track, you were always there with some scratch paper and tea. Time and again, you and Pamela graciously opened your home to me, and for that I am beyond grateful.

Thank you, Tamara Smyth and Charles Curtis, for your enthusiastic support. Thank you, Mauricio De Oliveira and Henry Abarbanel, for your thoughtful insights, and for agreeing to join this motley crew. It has been a pleasure working with each of you. I am privileged to be in such good company.

I would also like to thank my sister, Jessica, and my parents, Joe and Denise, for their patience and for providing emotional support when the pressure mounted. Even on Christmas.

Thank you, Adam Tinkle, for co-constructing so many successful performances with such an unlikely instrument. Let's do another West-Coast tour soon.

A huge number of people contributed in some capacity or another to the completion of this project, far too numerous to mention here. I am deeply grateful to them all.

Finally, I would like to thank the many stumbling blocks, hazards and pitfalls that reminded me that it's not enough simply to work hard. In order to survive we must adapt.

VITA

2007	B. A. in Music, Wesleyan University
2009	M.P.S. in Interactive Telecommunications, New York University
2018	Ph. D. in Music, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Nonlinear Modal Methods for Analysis and Synthesis

by

Joseph Mariglio

Doctor of Philosophy in Music

University of California San Diego, 2018

Professor Miller Puckette, Chair

An analysis and synthesis technique is described for the estimation and modeling of a particular class of nonlinear acoustical systems with energy-preserving nonlinearities. It is situated in the field of nonlinear sound synthesis techniques. The rich mathematical structure behind the technique is explored, with particular attention to the effects of scaling to arbitrarily high-dimensional systems. The technique is then applied to simulations and recordings of nonlinear vibratory behavior. After estimating model parameters from the timeseries, the models are validated using various assessment techniques, including resynthesis. Applications to computer music are discussed.

Chapter 1

Introduction

This publication documents an analysis and synthesis technique for the estimation and modeling of a particular class of nonlinear acoustical systems with losses in energy constrained to the linear dynamics. This particular class of system may be modeled as a multidimensional linear system perturbed by signal-dependent unitary transformations. It is situated in the field of sound analysis / synthesis techniques. The algebraic structure of the model affords an intuitively descriptive parameter-space, while allowing synthesis to guarantee stability, and analysis to converge on a solution with minimal complexity in the recursion. This thesis describes both the abstract mathematical properties of the underlying model as well as a family of algorithms for analysis and synthesis.

This introductory chapter first provides some context to the project, describing the nature of the recordings and performances that inspired the model this developed herein. Following this, Section 1.2 lays out our desired characteristics for this model and its parameter estimation algorithm, depicts the theoretical framework within which it resides, and makes note of some notational peculiarities developed for the specific needs of this project. Section 1.3 outlines the primary research questions this thesis hopes to answer, followed by an important caveat regarding equiavalency in the parameter space.

Finally, section 1.5 provides a brief guide to the rest of the document.

1.1 Performance

The theory for the model presented in later chapters largely grew out of informal observations and experiments with vibrometers such as piezoelectric contact microphones and hydrophones, electromagnetic pickups, and simple laser vibrometers. These forays were generally in the service of aesthetics or curiosity. Although these explorations may have lacked the rigor and procedure necessary to make definitive quantitative statements about the various systems that were observed, they nonetheless provided a qualitative and intuitive enrichment of understanding.

The subjects of investigation included systems such as spiderwebs in the presence of controlled jets of air, highly dispersive metal sheets, and the surfaces of various vibrating liquids. Their study comprised something more akin to a craft than a field of scientific inquiry. A few opportunities to perform with particularly beautiful systems in front of an audience presented themselves, and this modest performance practice necessitated more robust and portable realizations.

As a result, many iterations of vibrometers were developed over time. The larger goal was to create a portable array of such sensors, robust enough to perform outside of controlled conditions, yet still sensitive enough to deliver a satisfying musical experience. In addition to honing the design of the sensor, a digital signal processing algorithm was sought to compensate for unwanted components of the signal. Eventually, the more interesting questions came not from so-called "noise removal" techniques, but from parametric methods of modeling the system dynamics, based on recordings.

After several design iterations, these laser microphone arrays began to see utilization in musical performances, beginning sometime in 2014. Initially, these performances

had focused on spiderwebs. To activate the webs acoustically, a variety of mechanisms were tested. Ultimately it was determined that compressed air provided a highly controllable continuous activation not dissimilar to a bow. The results sounded like a cross between a string instrument and a reed instrument. Depending on the type of web, its size and age, these tiny instruments could have strings with fundamentals well below the audible range, and so audible activation would require coaxing out some upper vibratory mode. Some web configurations exhibited sympathetic resonance and coupling of an unusual character as the various modes interacted with each other along the diaphanous medium.

Although dexterity with the compressed air bowing technique improved with practice, the limitations of the medium became apparent rather quickly after a few performances. Most significantly, a reliable source of these instruments would require a reliable source of spiders, an animal which I find deeply and irrationally terrifying. This fact alone did not bode well for the future of the performance practice. Furthermore, while spiderwebs could occasionally be found in my apartment, their locations were very rarely convenient for performance. I had been lucky enough to find an orb web on a small potted orchid, and I had eked out as many performances and multi-channel recording sessions as possible with the instrument, but this web could not last forever. It turns out that spider silk dries out and becomes progressively inelastic as it ages. Finding another naturally occurring, highly portable, and –most importantly– *vacant* spider web could be months away. Even with this relatively convenient instrument, the process of calibrating the microphones was laborious, often requiring up to an hour of delicate, tedious work simply to prepare for a performance.

I eventually turned to other objects of interest, particularly those with properties uniquely suited to my laser microphone design. Liquids seemed especially promising. Minor modifications to the spiderweb technique afforded me the ability to amplify the

surface vibrations of a wide variety of liquids, foams and gels. Transducing signals into a dish of the liquid and listening to the result at the surface yielded audible transformations: a sort of electroacoustic filter. As might be expected, the resulting transformations were not strictly linear– differences in amplitude would reveal entirely different regimes of operation, and combinations of pure tones would occasionally interact in unexpected ways. I also found that feedback could be implemented by sending a combination of the microphone signals back into the transducer, resulting in "oscillators" with often unpredictable and rich behaviors.

These techniques in liquid amplification generated a number of performances in a wide variety of venues. Often, the sonifications were coupled with visualization techniques. In some, the light of an additional laser would be reflected off of a point on the surface of the liquid and onto a wall or projection screen in the performance space. The wave patterns along the surface of the liquid would then cause responsive, multifarious patterns in the laser projection. In others, an overhead projector was used to display the full image of activity on the surface of the liquid-air interface.

A number of performances, which were a collaborative effort with sound artist and multi-instrumentalist Adam Tinkle, integrated the dishes of amplified liquids with more traditional analog music and video synthesizer components. The source of the video was generally a small camcorder. A typical performance would begin with transducing sinusoids into the liquid– typically a solution of water and surfactant or potato flour– and transition into more complex processing relationships involving feedback. Simultaneously, a visual representation of the system was projected onto a wall or screen, and also routed to several television monitors via analog video mixer. Using video feedback both internally to this mixer and externally involving the camcorder, a variety of processing techniques, somewhat analogous to the sound processing, could take place in video streams. We experimented with software-controlled multicolor LEDs for additional stage

lighting and found, happily, that the PWM signals used to control these LEDs could be picked up by the laser microphones. I had experimented with techniques for reducing such interference during the microphone design stages, and ultimately decided to reduce my BOM by removing it altogether.

The gestalt of the performances was a combination of overwhelming physical presence of multi-sensory stimulation and garish equipment evoking the imagery of pseudoscience. In some performances we emphasized the latter of these by dressing in lab coats and delivering absurd abstracts from phony articles. The equipment was old and often failed in quirky, interesting ways. We destroyed at least three overhead projectors and burned out many more bulbs, each of which cost nearly as much as the cantankerous hulks themselves. It is difficult to assess the degree to which the synesthetic qualities of the performance would immerse our audiences, except to glean reactions afterward. My impression is that the overstimulating aspect of the performance was highly effective, given the oddball conversations that tended to happen as most people stumbled for the exits. Several of these performances have been documented with video, and a release of this material (in addition to more touring) is anticipated in future months.

1.1.1 Notes on the Recording Sessions

In addition to performances, this project generated a number of recording sessions, with varying degrees of control. One such recording session is described in detail in the following section. This thesis does not claim to have successfully modeled anything. Accomplishing that goal would have required an entirely different approach, with far more care taken to relate the model back to the particular system of interest. However, the fact remains that a framework for the analysis and synthesis of a particular class of nonlinear systems has been developed, and so have the measurement tools for exploring the systems that inspired this framework. An obvious way to proceed is to treat the

framework as a hypothesis, and attempt to prove it, but this is misguided. The framework is not a hypothesis, it is a sound analysis / synthesis technique. A more appropriate strategy might be to engage in a well-conducted recording session. The procedures bear some superficial similarity to each other, and there is strong justification for this: they both require precise control over environmental conditions. Regardless, the conclusions one can hope to draw from either form of inquiry are of firmly distinct categories.

1.1.1.1 Materials and Apparatus

A plastic container the shape of a rectangle with rounded edges (see figure 1.1), measuring $42.06748 \times 28.73248 \times 13.97$ cm was filled with 1 L of room temperature tap water. The sample was placed atop a 50 W, 8 Ω Vidsonix tactile transducer,

Figure 1.1: The apparatus, viewed from above.

driven by a Denon amplifier capable of supplying 80 W into 8 Ω with 0.08% THD, 100 dB SNR and ± 3 dB 20 Hz$\rightarrow$ 20 kHz response. The sample was coupled to the transducer using 4 neodymium magnets, each with 3.62874 kg pull strength. The sample was balanced carefully such that the water surface was roughly level across the entire container, resulting in a slightly varying depth of $1.3 - 1.7$ cm.

Two 5 mW red laser diodes were suspended above the sample using shock-absorbing mounts, and the light sources were received by a pair of high-bandwidth PIN photodiodes, forming a simple pair of geometric optical vibrometers. Vibrometer "A"

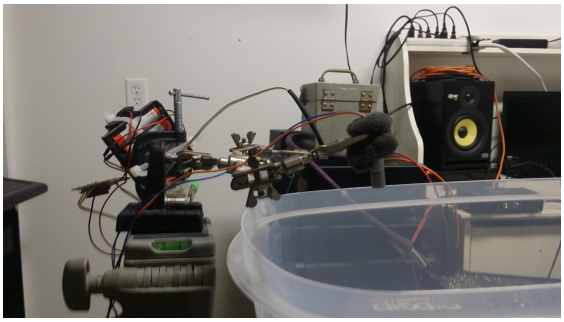

Figure 1.2: Vibrometer "A", measuring a point approximately 5 cm from the corner of the container.

was placed at a vertical distance of 12.5 cm from the surface of the water, with its receiver 14.5 cm above the surface. This pair measured a point approximately 5 cm from the corner of the container.

Vibrometer "B" was arranged such that the transmitter was 9.5 cm and the receiver 13.5 cm above the surface of the water. This pair measured a point approximately 14.366 cm from either side of the container (width-wise), and approximately 6 cm from the end of the container (length-wise).

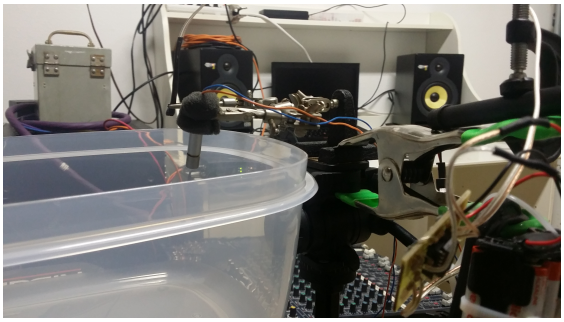

Figure 1.3: Vibrometer "B", measuring a point approximately 14.366 cm from either side of the container, 6 cm from the end.

The PIN photodiodes were configured in zero-biased photoconductive mode, amplified using a differential instrumentation transimpedance amplifier (TIA). [Gra96] The signals were then further amplified and their relative gains were calibrated using an analog Yamaha MG166c audio mixer and an oscilloscope. Ultimately the signals were

recorded at 48 kHz, 24 bit into a Sound Devices USBPre ADC.

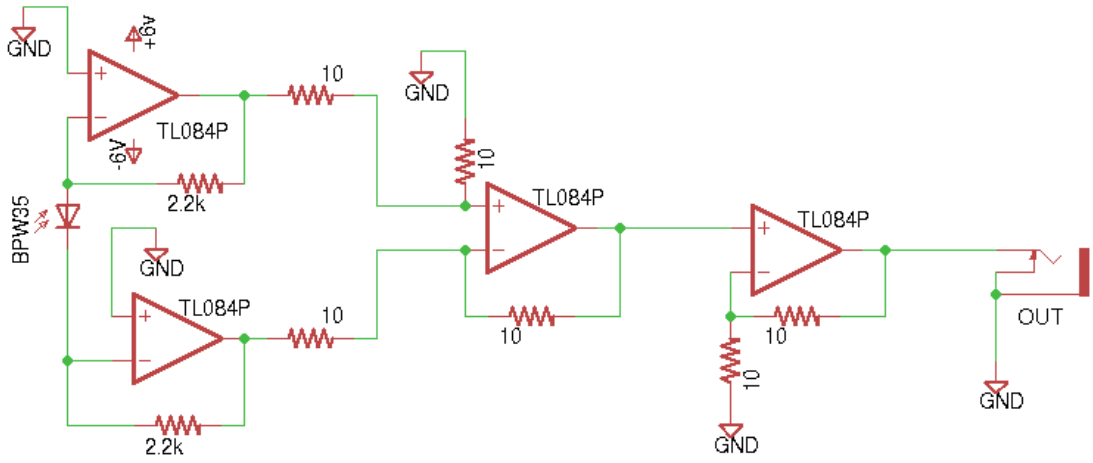

Figure 1.4: Schematic of the differential instrumentation TIA.

1.1.1.2 Methods

The sample was probed with a number of test signals, and the response of the sample recorded. These test signals were also digitally captured for reference. The signals consisted of white noise and linear sinusoidal chirps across several frequency ranges. The signal generation and capture was performed in the PureData audio synthesis environment. After each trial, the collected data was given an initial processing pass, whereupon spectrograms were generated to aid the search for interesting dynamics. Occasionally, based on the results of this search, follow-up tests were performed, prior to proceeding with the next trial.

1.1.1.2.1 Trial 1: Linear Chirps

After calibration, broad-band linear chirps were captured at 3 distinct amplitude levels, as described in the first 3 entries of table 1.1. After capturing the data, it was observed that the extreme low-end of the frequency range was the most active. Two

Table 1.1: A table of test signals used for Trial 1.

Trial No.	Time	Frequency Range	Amplitude
00.00	240s	0 Hz→ 24 kHz	0 dB
00.01	240s	0 Hz→ 24 kHz	−6 dB
00.02	240s	0 Hz→ 24 kHz	−12 dB
00.03	120s	0 Hz→ 100 Hz	0 dB
00.04	120s	0 Hz→ 50 Hz	0 dB

additional tests were performed within this range as a follow-up. The settings for these tests are also included in table 1.1, as entries 00.03 and 00.04.

1.1.1.2.2 Trial 2: White Noise

Table 1.2: A table of test signals used for Trial 2.

Trial No.	Time	Amplitude
01.00	120s	0 dB
01.01	120s	−6 dB
01.02	120s	−12 dB

White noise was then applied to the system, at 3 distinct amplitudes. Given the apparent low frequency poles that affected the system in Trial 1, we expected to see these poles respond to white noise. Furthermore, given what appeared to be amplitude-dependent sideband generation on those poles, we hypothesized that varying the amplitude of the white noise would result in varying the amplitude of those sidebands.

1.1.2 Notes

It is evident from the spectrograms that the two vibrometers measured very different transfer functions. In particular, vibrometer "B" appears to exhibit a considerable bias toward the low frequencies. Though this could have occurred for a number of reasons, the richness of high-frequency detail in the responses it measured in test 00.03 (figure A.4) and Test 00.04 (figure A.5) suggest this biasing occurred prior to the vibrometer receiver in the signal path.

The results of test 01-02b (figure A.8) are the most perplexing of this trial. In this test, the tank was vibrated with white noise at a relative amplitude of -12 dB. The resulting spectrogram shows some sort of amplitude modulation with an extremely low frequency mode. In fact, the modulation effects of this low frequency component can also be seen, albeit more subtly, in the -6 dB test (figure A.7). The darkened vertical bands that appear at regular intervals are evidence of this behavior's emergence. It appears that in a higher amplitude regime, the effect is less pronounced.

Examining the chirp responses in trial 1, we can see the dynamics are both nonlinear –due to the presence of additional striations to the initial chirp frequency– and scale-dependent –due to the response differences between tests with different amplitudes. Any modeling framework developed herein should be able to capture these dynamics.

1.2 Desiderata

We aim to model the behavior of systems similar to those described in the previous section. If we are successful, we will not only be able to describe systems in the wild via analysis, but also to model the behavior of entirely new systems. This possibility of synthesis has particular applicability to the field of computer music, which is enriched by the development of new synthesis paradigms. Given that we are interested in an analysis

/ synthesis algorithm pair, we can apply the following desiderata:

Firstly, we require minimal knowledge of the driving signal. In section 1.1.1.2, we used two driving signals which are common to the field of system identification. Their pervasive adoption is justified: using these driving signals, we know we can completely identify the dynamics of a linear system. That being said, the chirps used in test 1 are unlikely to resemble much of anything we will find in the wild, and certainly far more specialized than the mean-zero white noise used in test 2. We therefore would prefer, to whatever extent possible, that the driving function be as simple and unconstrained as possible. We go so far as to specify that it should be a random variable of some sort.

Secondly, we would like to end up with an interpretable parameter space which can be inspected to observe qualities like the proximity (or in the extreme, equivalence) of one system to another in terms of dynamics. An even happier result would be that the parameter space may give us some information about the time scale of the system dynamics, losses due to imperfect conservation of energy, et cetera. For the purposes of a musical synthesis algorithm, we might ask for a predictable set of center frequencies, bandwidths, and some classification or quantification of timbral complexity.

Third, we would like as simple a parameter set as needed to explain the system dynamics up to a tolerance. It is common in system identification to require the simplest explanatory model conceivable, thereby mitigating the hazards of overfit. The more complex a model is, the greater these hazards, therefore this requirement is roughly equivalent to prioritizing solutions which are more likely to generalize well.

1.2.1 Framework

The standard framework for modeling a general system is presented in this section. We include this primarily as an opportunity to demonstrate the canonical notational conventions and explain the necessity of extending them for the purpose of

this dissertation.

1.2.1.1 The Residual

Recall that a linear model is defined as a system of equations whose future output may be encapsulated by a linear combination of previous inputs and outputs, i.e.:

$$\boldsymbol{y}_k = \Theta \boldsymbol{p}_k + \boldsymbol{\xi m}_k \tag{1.1}$$

where $\boldsymbol{p}$ is a vector of the previous input and output terms, and Θ is a parameter matrix describing how these are combined to describe each successive timestep. Such a model is generally considered incomplete without $\boldsymbol{\xi m}$, which models the non-deterministic component of the measurement. In modeling, we are attempting to find the best approximations of the above terms and parameters, $\hat{\Theta}$, $\hat{\boldsymbol{\xi m}}$, and, in the case of prediction, $\hat{\boldsymbol{y}}$:

$$\hat{\boldsymbol{y}}_k = \hat{\Theta} \boldsymbol{p}_k + \hat{\boldsymbol{\xi m}}_k \tag{1.2}$$

Note in equation (1.2) we imply that the term vector $\boldsymbol{p}$ is directly observable and not approximate. Eventually we may in fact find approximations of past terms or even future ones. This becomes significant in recursive algorithms, which tend to alternate between prediction and correction steps.

The *Linear Residual* is simply the point-by-point difference between this linear approximation and the measured output of the system, disregarding measurement noise, which will be addressed in chapter 3:

$$\boldsymbol{e}_{\text{lin}k} = \boldsymbol{Y}_k - \hat{\boldsymbol{A}} \boldsymbol{Y}_{k-1} \tag{1.3}$$

where $\hat{\boldsymbol{A}}$ is simply the estimated parameter matrix whose entries only pertain to

previous system outputs. Note that the system described in equation (1.3) is linear with respect to measurements because its term matrix $\boldsymbol{Y}$ is always constrained to past output values. This system would remain linear with respect to the measurements if the model included input terms or linear transformations of input or output terms. In the linear case, numerous techniques, such as those presented in [Alp14] , [Nor86, p. 59] , [Abu12, p. 82] , [LS86, p. 33] and [Bil13, p. 20], and many others, may be used to identify the system dynamics. These classical identification techniques, which include LPC, modal analysis and Kalman filtering, form the ground from which linear system identification has been cultivated.

A nonlinear model, by extension, uses function composition rather than matrix multiplication to express its transition from one timestep to the next. Despite this, it can be represented in much the same form as equation (1.1), by relaxing the constraint on $\boldsymbol{P}$ to include nonlinear functions of input or output terms. NARMAX, a deterministic treatment of this generalization, may be found in [Bil13, p. 36]. We will explore these ideas in greater depth, establishing our algorithm in relation to them, in chapters 2 and 3.

1.2.1.2 Extended Notation

In the literature, it is common to use the above notation for tensors without regard to which reference frame we are viewing the tensor from. Typically this is permissible because there are relatively few bases which we are operating in to solve a particular problem, and once that basis is determined, we need not concern ourselves with further transformations until we actually need to pull coefficients out of the tensors. At that time, these coefficients must be represented in a particular basis. However, this dissertation makes use of multiple bases, among which we will need to transform to solve different pieces of the problem. Due to the fact that we relax the linearity requirement on the

system, we lose the privilege of contracting our change-of-basis operations into a single transformation step, and therefore we must have a consistent notation for a given tensor under a certain set of transformations.

To denote a particular basis under which we view a given tensor, we place the chain of transformations required to get back to standard basis over the top of the tensor, like so:

$$\begin{aligned} \overset{\boldsymbol{SE}}{\boldsymbol{x}} &\equiv \boldsymbol{x} \\ \overset{\boldsymbol{SE}}{\boldsymbol{T}}\boldsymbol{E}^{-1}\boldsymbol{S}^{-1} &\equiv \boldsymbol{T} \end{aligned} \tag{1.4}$$

where $\boldsymbol{S}$ is a matrix which diagonalizes the column state-variable vector $\boldsymbol{x}$ and state-transition matrix $\boldsymbol{T}$, and $\boldsymbol{E}$ is a matrix which transforms the diagonalized tensors into the real-valued *energy basis*. This basis will allow us to track the amount and directionality of energy flows among modes of vibration, which under the linear approximation are presumed to be independent. We will assume that conservation of energy will provide us with a dimensionally-reduced representation of system dynamics, thereby reducing the complexity of the calculations required to solve for higher degree nonlinearities. For example, solving a quadratic system with 4 degrees of freedom would require solving for a tensor containing 64 terms, which, using this dimensonality reduction strategy, we can shrink to a 20-dimensional subspace. This dimensionality reduction step itself, along with the fact that it is *necessary* for solving this problem is a central finding of this dissertation. We define it in Section 3.3.3.

1.2.2 Two Cross-Model Metrics

To help us on this odyssey, we will need two metrics that may be considered applicable in a cross-model sense: one which pertains to the modeling side of the problem, and another which pertains to the identification side. In the case of modeling, we might be in a position to use a metric as an equivalence condition or distance criterion among candidate models. This might be useful in cases where dynamics are recorded, and therefore we only have access to approximations of state. It might also be useful if the system in question is synthesized, and therefore the state is known as ground truth. In the case of identification, we may find use for a recursive approach, and this requires a metric to determine if a subsequent recursion of the identification represents a step toward or away from the solution.

1.2.2.1 Index of Nonlinearity

The task of comparing two nonlinear systems, even if their parameters are known *a priori*, is somewhat less straightforward than that of comparing two linear systems, for several important reasons. First, and perhaps most obviously, the nonlinear component of the model must be somehow disentangled from the linear component, and a new metric comparing the two nonlinear components must be constructed. (Linear systems are somewhat more simple to compare, via their eigendecompositions.) Second, this nonlinear component is no longer guaranteed to be scale-invariant, a degree of freedom that threatens, among other things, to obfuscate the dynamics at the source with characteristics typically attributed to the measurement procedure. As we will see in section 4.1.1.2, this nonlinear component might vary non-uniformly along its dimensions. For these reasons and others, we require the definition of an intrinsic "Nonlinearity Index" which will serve as a metric in the space of nonlinear systems. We introduce this measure in earnest in section 3.3.5.2. This will serve as an equivalence condition between models

based on their overall dynamics, and may be calculated only if the parameters of the system are known. Of course, that will not prevent us from attempting to estimate it. Such a measure captures the degree to which the system deviates from linearity, independent of scale, modeling error, measurement noise, and linear frequency.

1.2.2.2 Absolute Prediction Error

Concomitant with estimating the parameters of a system which might be either predominantly linear or nonlinear, there is a hazard of selecting an inappropriate model to fit the in-sample data. This sort of error invalidates the parameter estimation and causes poor generalization out-of-sample. When selecting a model, the general rule of thumb is to start with the simpler model and build upward in complexity. Implicitly, two descriptions being equally accurate, we often prefer the simpler of the two. To facilitate this, we must construct a metric which encapsulates the improvement in descriptive ability between echelons of complexity. In our case, the degrees of a Taylor series provides us with a convenient hierarchy of analyses. We describe such a measure, which is itself invariant to other model parameters like scale, etc, in section 3.3.5.1.

1.3 Research Questions

We set out to answer a number of questions regarding general dynamical systems, pertaining to the problem of modeling dynamics given measurements of system output alone. Some of these questions will be answered in the process of designing our conjoint synthesis and analysis algorithms, and some of these questions must be answered in a separate process of assessment given large numbers of systems. Since in an epistemological sense, we cannot know the state variables underlying a given system (state itself being a conceptual tool rather than an intrinsic property of processes), we rely on

synthesized systems of a particular form to pit our algorithms against. This permits us to vary specific classes of parameters while keeping other parameter classes constant, affording us the ability to validate the algorithms along a variety of axes. The questions we ask are therefore related to our ability to perform this assessment: whether given parameter sets will converge under estimation, whether systems of given parameter sets are equivalent or nearly so, and what the expected accuracy of a given estimation might be.

We have stated that we would like for our analysis algorithm to make very few assumptions about the system to be identified. Since we will be relying mostly on synthesized systems, we are afforded the opportunity to investigate how the presence or absence of *a priori* knowledge of the system affects our ability to identify it. In chapter 4, we organize our tests for the comparison of three basic cases: knowledge of the spatial arrangement of microphones, blind estimation, and knowledge of the linear dynamics. Of these three, we anticipate the second case to be the most challenging.

1.3.1 Bandwidth of Measurements

Our measurement techniques provide us with a finite-bandwidth, discrete-time representation of the signals, and our resulting models of the underlying system share those constraints. What is the effect of increasing the frequency with which the system is measured on our ability to model it? Our choice of the Taylor series expansion suggests that as our model's sampling frequency increases, our low-order approximation will approach the underlying measured system. (C.f. Section 3.2.1.) Is increasing the sampling frequency similar to increasing the dimensionality of the measurements, from the perspective of accuracy? We intuit that these two factors might have a similar effect on the analysis algorithm, in that we increase the overall data available to the algorithm in both cases. How might these two dimensions differ?

1.3.2 Window Size

Another important factor in our analysis is the length of time needed to identify underlying system dynamics. Intuitively, we hypothesize that a longer window size would correspond with a greater degree of accuracy, perhaps at the cost of computational complexity. We will seek to explore both sides of this implied trade-off. Is there any additional cost implied by increasing this window size? In Fourier analysis, particularly in STFT applications such as phase vocoders, there is one such famous compromise between time- and frequency-domain acuity for the detection of energy packets. Will our analysis algorithm suffer from similar constraints?

1.3.3 Dimensionality of Measurements

An early intuition held that using more microphones might provide us with more reliable measurements, regardless of how those microphones might be arranged in space. What is the effect of adding more dimensions to our measurements? How does the algorithm for analysis scale with measurement vector dimensionality? In section 1.2.1.2, we hinted at the possibility that the solutions we will seek lie within a subspace of the state-space as it is presented to us by more conventional means, and stated that dimensional reduction comprises an essential tool which we wield in order to solve the problem in general. This process naturally leads to the partitioning of the larger problem space into multiple subspaces. How do these component subspaces of the algorithm scale with the dimensions of the measurement vector? Is there an optimal dimensionality which our measurements must have to assure a particular level of accuracy or convergence? If we assume the system is autonomously being excited by, e.g. ambient noise, what is the dimensionality of that noise? What happens if this excitatory process noise lies in a subspace of the measured system?

1.4 A Note on Equivalence

Since we can not observe state directly in the wild, we must recall that any transformation of state parameters might yield an observably identical– and thus equally valid– estimated system, assuming, of course, that this transformation is carried properly throughout the description of the system. Therefore, in our parameter estimation, we seek to locate the system within an equivalence class. In the linear case, this equivalency may be represented as a transformation of coordinates. How does this change for general systems? How might the inherent uncertainty about the representation of system state variables interfere with our ability to find nonlinear relationships between those state variables to explain higher-degree nonlinear dynamics? Can this uncertainty be dealt with in a practical way as we search for optimally explanatory variables?

1.5 Structure of the Dissertation

The derivation of the model begins in earnest in chapter 2. Following this description of the synthesis technique, chapter 3 explains the analysis algorithm. Chapter 4 catalogs a number of test batteries designed to validate the analysis / synthesis pair, directly responding to the questions posed in section 1.3. Finally, chapter 5 offers some high-level assessment and documents future work.

Chapter 2

General Model Derivation (Synthesis)

This chapter develops a nonlinear model with a rich internal structure, suitable for describing a wide variety of nonlinear dynamics. We first derive the framework for the model as a perturbation on linear systems. We then construct the model from its simplest, most constrained form to its most general, observing how the equations increase in dimensionality while preserving their structure.

2.1 Linear Systems

In this section, we briefly review topics related to linear models, with the intention of establishing a notational convention. These basic concepts will provide the foundation on top of which we will construct our nonlinear model in section 2.2.

2.1.1 Input-Output Representation

We begin by considering a linear model of the form:

$$A(z^{-1})y(k) = \frac{B(z^{-1})}{F(z^{-1})}u(k) + \frac{\gamma(z^{-1})}{D(z^{-1})}\xi(k) \tag{2.1}$$

where $k \in \mathbb{Z} : y(k)$ is the output signal, $u(k)$ is the input signal, and $\xi(k)$ is an independent and identically distributed noise sequence with expected value $E(\xi) = 0$ and finite variance. [Bil13, p.18] This representation is generally called an *Input-Output* model. [LS86, p.419] In this notation, the term z^{-1} signifies a backward shift on a signal, i.e. $z^{-1}p(n) = p(n-1)$ for all signals p. The *model parameter* sequences

$$
\begin{aligned}
A(z^{-1}) &= \sum_{n=0}^{Na} a_n z^{-n} = 1 + a_1 z^{-1} + \cdots + a_{Na} z^{-Na} \\
B(z^{-1}) &= \sum_{n=1}^{Nb} b z^{-n} = b z^{-1} + \cdots + b_{Nb} z^{-Nb} \\
\gamma(z^{-1}) &= \sum_{n=0}^{N\gamma} \gamma_n z^{-n} = 1 + \gamma_1 z^{-1} + \cdots + \gamma_{N\gamma} z^{-N\gamma} \\
D(z^{-1}) &= \sum_{n=0}^{Nd} d_n z^{-n} = 1 + d_1 z^{-1} + \cdots + d_{Nd} z^{-Nd} \\
F(z^{-1}) &= \sum_{n=0}^{Nf} f_n z^{-n} = 1 + f_1 z^{-1} + \cdots + f_{Nf} z^{-Nf}
\end{aligned}
\tag{2.2}
$$

describe the behavior of the system as combinations of the *model term* signals $y(k)$, $u(k)$, and $\xi(k)$. When $z = e^{i\omega T}$, angular frequency ω is in the range of $-\pi < \omega \leq \pi$, and T is the sampling interval, the model can be described in the frequency domain. The stability of the system is entirely determined by the roots of the model polynomials $A(z^{-1}), D(z^{-1}), F(z^{-1})$, assuming these share no common factors with the polynomials $\gamma(z^{-1})$ and $B(z^{-1})$. [Smi07, p.127] [Bil13, pp.18-19] In general, an oscillating system will have complex roots expressible in polar form as $\rho e^{i\omega T}$. For our purposes, we will simply define the polynomials $F, D \equiv 1$. Note that the number of roots the model exhibits is determined by the degree of each of the polynomials: $Na, Nb, N\gamma, Nd, Nf$. This general model can be expanded from the single-input, single-output case (SISO) to cases with multiple-inputs and multiple-outputs (MIMO), with careful attention to the effects on

model complexity and stability. [Smi07, p.86]

In many cases, some of the model terms cannot be directly observed. Depending on the system in question, model terms may be recast as random variables, or removed from the system entirely. In this case, $\xi(k)$ absorbs the statistical properties of the unknown variable. For example, if the input is unknown, but can be assumed to be mean-zero and of finite variance, equation (2.1) might become

$$A(z^{-1})y(k) = \gamma(z^{-1})\xi(k), \tag{2.3}$$

effectively by setting the model parameters $B(z^{-1}) = 0$ and $D(z^{-1}) = 1$, and describing the system "input" as a random variable ξ. The model described by equation (2.3) is commonly known as *Autoregressive Moving Average* or *ARMA*, and is quite popular in modeling timeseries assumed to have linear dynamics. [Bil13, p.19]

We can rearrange the terms in equation (2.3) to place the current measured output of the system $y(k)$ as a combination of previous outputs and noise, by negating the output model parameters, yielding the difference equation for an *infinite impulse response (IIR)* filter: [Smi07, p.48]

$$y(k) = -A(z^{-1})y(k-1) + \gamma(z^{-1})\xi(k). \tag{2.4}$$

Note that the zeroth term of $A(z^{-1})$, $a_0 z^0 = a_0$, is a gain factor which we omit for clarity, essentially assuming $a_0 = 1$. If there is a non-unity coefficient here, it encodes the overall gain characteristic of the difference equation, and can be absorbed into the right-hand side of the equation algebraically.

In this case we continue with the assumption that we do not know the input signal, so we model it as a stochastic process. However, we could have also modeled the input deterministically, resulting in a *Autoregressive Moving Average with eXogenous inputs*

(*ARMAX*) model. In this case, equation (2.4) becomes an IIR filter with a noise term:

$$y(k) = -A(z^{-1})y(k-1) + B(z^{-1})u(k) + \gamma(z^{-1})\xi(k) \tag{2.5}$$

It is also trivial to derive the deterministic input-output transfer function of the filter as the ratio of the two polynomials in z^{-1}: [Smi07, p.121]

$$H(z) = \frac{B(z^{-1})}{A(z^{-1})} = \frac{\sum_{n=0}^{N_b} b_n z^{-n}}{\sum_{n=0}^{N_a} a_n z^{-n}} = \frac{1 + b_1 z^{-1} + \cdots + b_{N_b} z^{-N_b}}{1 + a_1 z^{-1} + \cdots + a_{N_a} z^{-N_a}} \tag{2.6}$$

and the stochastic input-output transfer function in the same way:

$$H(z) = \frac{\gamma(z^{-1})}{A(z^{-1})} = \frac{\sum_{n=0}^{N_\gamma} \gamma_n z^{-n}}{\sum_{n=0}^{N_a} a_n z^{-n}} = \frac{1 + \gamma_1 z^{-1} + \cdots + \gamma_{N_\gamma} z^{-N_\gamma}}{1 + a_1 z^{-1} + \cdots + a_{N_a} z^{-N_a}} \tag{2.7}$$

The roots of $A(z)$ comprise the poles of the transfer function, while those of $\gamma(z)$ and $B(z)$ comprise the zeros. One of the advantages of assuming a linear model is the ease with which one may convert between these and other equivalent representations of the system. [Bil13, p.149] Another is the ease with which a relationship between models of the same class might be computed. [Bil13, p.33]

We can write equation (2.5), with no loss of generality, as

$$y_k = \boldsymbol{p}_k \theta + e_k, \tag{2.8}$$

simply by concatenating all the model terms– not including the current measurement y_k, but certainly including lagged output terms like y_{k-1} if they are significant– into the row vector $\boldsymbol{p}_k$, and all the model parameters into the N_θ-dimensional column vector θ. For notational convenience we have encoded our model term sequences as K-dimensional vectors $\boldsymbol{y}$ and polynomials $A \ldots F$ as vectors of their coefficients. Note that in order for

the multiplication to be defined, the row-vector $\boldsymbol{p}_k$ must be $N\theta$-dimensional. [Bil13, p.65] In this broad view, we see the model as a linear transformation of model terms.

The residual term e_k is the generalized error in the model's prediction. Note that $\boldsymbol{e}$ may in some cases be equivalent to $\boldsymbol{\xi}$ or $\boldsymbol{u}$, depending on the particulars of the model. However, even when we place $\boldsymbol{\xi}$ or $\boldsymbol{u}$ into our model term vector, we must always account for $\boldsymbol{e}$, which then describes any deficiencies in the model.[Nor86, p.60] The analysis of $\boldsymbol{e}$ is extremely important in solutions to the model-fitting problem, discussed further in section 3.3.1.

2.1.2 Similarity Transformations and State-Space Representation

Any system that can be described in *Input-Output* form may alternately be described in its *State-Space* form. [LS86, p.419] By invoking such a representation, we acknowledge that the system behavior could be determined by some internal state, which we may not measure directly. In our case this representation will be dependent on a continuously varying term x, which enters the system equations as follows:

$$\begin{aligned} \boldsymbol{x}_k &= \boldsymbol{T}\boldsymbol{x}_{k-1} + \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{C}\boldsymbol{\xi}\boldsymbol{p}_k \\ \boldsymbol{y}_k &= \boldsymbol{W}\boldsymbol{x}_k + \boldsymbol{\xi}\boldsymbol{m}_k \end{aligned} \tag{2.9}$$

The system described above implies a linear relationship between measured output and internal state. However, the concept of state may be further generalized to nonlinear relationships as well, using functional composition instead of matrix multiplication, to transform from state space to measurement space.

2.1.2.1 Similarity Transformations

It will occasionally be advantageous to inspect or construct our model using various coordinates so that certain features may become apparent or computations may be made more efficiently. For example, we might want to change the appearance of the state transformation matrix $\boldsymbol{T}$, without changing the linear operator it represents. Such a change in coordinates is one-to-one and onto, and therefore invertible, and is called a *similarity transformation*. We will use this particular technique so frequently that we introduce a shorthand notation for it. Similarity transformations can be performed on matrices representing linear operators by matrix multiplications of the form $\overset{\boldsymbol{S}}{\boldsymbol{T}} \equiv \boldsymbol{S}^{-1}\boldsymbol{T}\boldsymbol{S}$. The related operation to change the coordinates of a column vector is simply a linear transformation: $\overset{\boldsymbol{S}}{\boldsymbol{x}} \equiv \boldsymbol{S}^{-1}\boldsymbol{x}$. In general, the application of a similarity transformation on a (p,q)-tensor requires a forward multiplication on the q covariant axes and a multiplication by its inverse on the p contravariant axes. Similarity transformations allow us to separate the underlying operator from its representation in a particular basis. [SB97, p.227] This notation is an extension and specialization of the more common convention of applying a diacritic to refer to a vector, matrix, or tensor in a different space. The definitions of system terms and parameters using this shorthand will always be explicitly defined in the context of their derivation and usage. While this custom of explicitly defining the transformation rules for each object may seem redundant at first, in later derivations the practice become invaluable to a principled understanding of the problem.

The state-space model in equation (2.9) can therefore be represented by any linear transformation $\boldsymbol{S}$ of the state vector $\overset{\boldsymbol{S}}{\boldsymbol{x}}$, provided the change of basis is consistently applied. This yields the form for the state-space model under a similarity transformation

into a space determined by the invertible matrix $\boldsymbol{S}$:

$$\begin{aligned}
\overset{\boldsymbol{S}}{\boldsymbol{x}} &\equiv \boldsymbol{S}^{-1}\boldsymbol{x} \\
\overset{\boldsymbol{S}}{\boldsymbol{x}}_k &= \boldsymbol{S}^{-1}\boldsymbol{T}\boldsymbol{S}\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1} + \boldsymbol{S}^{-1}\boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{S}^{-1}\boldsymbol{C}\boldsymbol{\xi}\boldsymbol{p}_k \\
\boldsymbol{y}_k &= \boldsymbol{W}\boldsymbol{S}\overset{\boldsymbol{S}}{\boldsymbol{x}}_k + \boldsymbol{\xi}\boldsymbol{m}_k
\end{aligned} \tag{2.10}$$

Applying this change of basis to $\boldsymbol{x}$ allows us to derive a representation of the entire system viewed in this basis:

$$\begin{aligned}
\overset{\boldsymbol{S}}{\boldsymbol{T}} &\equiv \boldsymbol{S}^{-1}\boldsymbol{T}\boldsymbol{S} \\
\boldsymbol{\Lambda} &\equiv \overset{\boldsymbol{S}}{\boldsymbol{T}} \\
\overset{\boldsymbol{S}}{\boldsymbol{B}} &\equiv \boldsymbol{S}^{-1}\boldsymbol{B} \\
\overset{\boldsymbol{S}}{\boldsymbol{C}} &\equiv \boldsymbol{S}^{-1}\boldsymbol{C} \\
\overset{\boldsymbol{S}}{\boldsymbol{W}} &\equiv \boldsymbol{W}\boldsymbol{S}
\end{aligned} \tag{2.11}$$

Substituting these transformed terms and parameters into equation (2.10) demonstrates the visual clarity this notational choice affords us:

$$\begin{aligned}
\overset{\boldsymbol{S}}{\boldsymbol{x}}_k &= \boldsymbol{\Lambda}\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1} + \overset{\boldsymbol{S}}{\boldsymbol{B}}\boldsymbol{u}_k + \overset{\boldsymbol{S}}{\boldsymbol{C}}\boldsymbol{\xi}\boldsymbol{p}_k \\
\boldsymbol{y}_k &= \overset{\boldsymbol{S}}{\boldsymbol{W}}\overset{\boldsymbol{S}}{\boldsymbol{x}}_k + \boldsymbol{\xi}\boldsymbol{m}_k
\end{aligned} \tag{2.12}$$

2.1.2.2 The Modal Basis

A particularly elucidating view of a linear transformation $\boldsymbol{T}$ is achieved when the the similarity transformation is defined by an eigenvector basis $\boldsymbol{S}$ of $\boldsymbol{T}$. [Smi07,

p.363] If the eigenvectors of $\boldsymbol{T}$ are linearly independent, then $\boldsymbol{S}$ diagonalizes $\boldsymbol{T}$. This represents the action of the linear operator encoded by $\boldsymbol{T}$ as a collection of the most elementary operations: a multiplication by a scalar. [Str09, p.358] Representing the dynamics in such a way permits us to decouple the actions of each degree of freedom in the system. This diagonal matrix $\overset{\boldsymbol{S}}{\boldsymbol{T}}$ is so important to the study of systems that we give it its own variable: $\boldsymbol{\Lambda}$. The nonzero entries of this matrix are the eigenvalues of the linear transformation. A synonym for this diagonalized state space view of a system is its *modal representation*.[Smi07, p.362] Explicitly writing out some of the entries in the state transition equations may demonstrate why this is such an important representation:

$$\begin{pmatrix} \overset{\boldsymbol{S}}{x}_{1,k} \\ x_{2,k} \\ x_{3,k} \\ \vdots \\ x_{N\theta,k} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ \vdots & \ddots & \lambda_3 & \ddots & \vdots \\ 0 & \dots & 0 & \ddots & 0 \\ 0 & 0 & \dots & 0 & \lambda_{N\theta} \end{pmatrix} \begin{pmatrix} \overset{\boldsymbol{S}}{x}_{1,k-1} \\ x_{2,k-1} \\ x_{3,k-1} \\ \vdots \\ x_{N\theta,k-1} \end{pmatrix} + \begin{pmatrix} \overset{\boldsymbol{S}}{b}_1 \\ b_2 \\ b_3 \\ \vdots \\ b_{N\theta} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} u_{k-0} \\ u_{k-1} \\ u_{k-2} \\ \vdots \\ u_{k-N\theta} \end{pmatrix} + \begin{pmatrix} \overset{\boldsymbol{S}}{c}_1 \\ c_2 \\ c_3 \\ \vdots \\ c_N \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \xi p_{1,k-1} \\ \xi p_{2,k-1} \\ \xi p_{3,k-1} \\ \vdots \\ \xi p_{N\theta,k-1} \end{pmatrix}$$

$$\overset{\boldsymbol{S}}{\boldsymbol{x}}_k = \boldsymbol{\Lambda}\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1} + \overset{\boldsymbol{S}}{\boldsymbol{b}}\boldsymbol{u}_k + \overset{\boldsymbol{S}}{\boldsymbol{c}}\boldsymbol{\xi}\boldsymbol{p}_{k-1} \tag{2.13}$$

In the $\boldsymbol{\Lambda}\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1}$ term, every coefficient in $\boldsymbol{\Lambda}$ multiplies only one state variable in $\overset{\boldsymbol{S}}{\boldsymbol{x}}$. The eigenvalues entirely determine the locations of the poles in the system transfer function, and furthermore, they do so independently. This is analogous to a model where each individual feedback path is applied in parallel (in the process update step) and finally mixed together (in the output step). Thus a modal decomposition is a generalization of the partial fraction expansion of a filter. [Smi07, p.363] The terms $\boldsymbol{S}^{-1}\boldsymbol{b}\boldsymbol{u}_k$ and $\boldsymbol{S}^{-1}\boldsymbol{c}\boldsymbol{\xi}\boldsymbol{p}_{k-1}$ encode how the input signal and process noise couple to each of the modes.

In the case of linear systems, it is important to note that any matrix of linearly-independent eigenvectors of $\boldsymbol{T}$ forms a basis for the modal space of $\boldsymbol{T}$. Each of the

columns may be scaled by any complex number, or the order of the columns permuted, and the resulting matrix is still a basis for the modal space. We have the same freedom of choice for the ordering of the diagonal elements in $\mathbf{\Lambda}$ as well, as long as we consistently apply the same permutation to the eigenvector columns as we did to the diagonal coefficients in $\mathbf{\Lambda}$. To form an ordered basis, we therefore apply the convention such that the eigenvalues are sorted in descending order, according to their real component first, the absolute value of their imaginary component second, and the sign of the imaginary component third. This places conjugate modes next to each other, and orders them according to their real components, e.g.:

$$\mathbf{\Lambda} \equiv \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \bar{\lambda}_1 & 0 & \dots & 0 \\ \vdots & \ddots & \lambda_3 & \ddots & \vdots \\ 0 & \dots & 0 & \ddots & 0 \\ 0 & 0 & \dots & 0 & \bar{\lambda}_{N\theta-1} \end{pmatrix} \quad \& \quad \boldsymbol{S} \equiv \begin{pmatrix} s_{1,1} & \bar{s}_{1,1} & s_{1,3} & \dots & \bar{s}_{1,N\theta-1} \\ s_{2,1} & \bar{s}_{2,1} & s_{2,3} & \dots & \bar{s}_{2,N\theta-1} \\ \vdots & \vdots & \vdots & \vdots & \\ s_{N\theta,1} & \bar{s}_{N\theta,1} & s_{N\theta,3} & \dots & \bar{s}_{N\theta,N\theta-1} \end{pmatrix} \tag{2.14}$$

The above definitions show the equivalences among the coefficients for clarity. We further specify that $\boldsymbol{S}$ must be normalized such that its largest singular value is 1 and its last row $(s_{N,1}, \dots, s_{N,N})$ lies entirely on the real axis. Now our definition of $\boldsymbol{S}$ and $\mathbf{\Lambda}$ is sufficiently specific that these objects are unique given a set of eigenvectors.

For oscillating systems, the values of the diagonal entries in $\mathbf{\Lambda}$ are in $\mathbb{C}$. The actions of these entries can best be understood in polar coordinates. Let ω_n be the angle of the complex eigenvalue λ_n, and ρ_n be its radius. The response of each eigenvalue can

be expressed in the frequency domain as:

$$\begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_N \end{pmatrix} = \begin{pmatrix} \rho e^{i\omega_1 T} & 0 & \dots & 0 \\ 0 & \rho e^{i\omega_2 T} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \rho e^{i\omega_N T} \end{pmatrix} \tag{2.15}$$

The eigenvalues in $\mathbf{\Lambda}$ perform an analogous role to the roots of the polynomial A in equation (2.1). The parameters in both representations perform the action of infinitesimal rotation and scaling necessary for the discrete-time description of vibrating systems. In both cases, the polar representation expresses the frequency of the pole to be $\frac{\omega_n}{2\pi T}$ and the pole's "Q-Factor" to be roughly $\frac{1}{1-\rho_n}$, for points close to the unit circle.

2.1.2.3 Similarity Transformation into $\mathbb{R}^N$

Another useful similarity transformation is the one which transforms the complex-valued, pairwise conjugate modal basis described in section 2.1.2.2 into a block-diagonal, real-valued modal basis. The transformation will expand the diagonal complex-conjugate eigenvalue pairs in $\mathbf{\Lambda}$ into 2×2 rotation matrices $\boldsymbol{R}$. We will refer to 2-blocks in this coordinate system as *real modes*, which consist of conjugate pairs of complex states in the $\boldsymbol{S}$ basis.

$$\begin{aligned} \boldsymbol{E}^* &\equiv \boldsymbol{E}^{-1} \\ \boldsymbol{R} &\equiv \boldsymbol{E}^* \boldsymbol{S}^{-1} \boldsymbol{T} \boldsymbol{S} \boldsymbol{E} \\ \boldsymbol{x}_{k-1} &\equiv \boldsymbol{S}\boldsymbol{E}\boldsymbol{x}^{\boldsymbol{SE}}_{k-1} \end{aligned} \tag{2.16}$$

We choose a unitary basis $\boldsymbol{E}$ to accomplish this change of coordinates, which we define to be:

$$\boldsymbol{E}(2) \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \tag{2.17}$$

in the two-state case, and with the block structure:

$$\boldsymbol{E}(N) \equiv \begin{pmatrix} \boldsymbol{E}(2) & & \\ & \ddots & \\ & & \boldsymbol{E}(2) \end{pmatrix} \tag{2.18}$$

in the N-state case. Recall that we are only adding rows and columns in pairs. We implicitly specify that the blocks of rotation matrix $\boldsymbol{R}$ perform a clockwise rotation.

2.1.2.4 Controllability and Observability

In the interest of making empirical identification possible, we apply the assumption that the system we are modeling is fully *observable* at its outputs $\boldsymbol{y}$. Furthermore, because we would prefer not to use measurement techniques that require us to drive the system with a particular input, we also assume the system is fully *controllable* at its deterministic inputs $\boldsymbol{u}$ and stochastic inputs $\boldsymbol{\xi}$. For the system to be observable at a given output, the following row vectors must form an invertible matrix:

$$\begin{bmatrix} \begin{pmatrix} \boldsymbol{w}_{1,n} & \dots & \boldsymbol{w}_{m,N} \end{pmatrix} & \boldsymbol{T}^0 \\ \begin{pmatrix} \boldsymbol{w}_{1,n} & \dots & \boldsymbol{w}_{m,N} \end{pmatrix} & \boldsymbol{T}^1 \\ & \vdots \\ \begin{pmatrix} \boldsymbol{w}_{1,n} & \dots & \boldsymbol{w}_{m,N} \end{pmatrix} & \boldsymbol{T}^N \end{bmatrix} \tag{2.19}$$

Similarly, in order for the system to be controllable at a given input, the following column vectors must form an invertible matrix for the column vector of $\boldsymbol{B}$ corresponding to that input:

$$\left[\boldsymbol{T}^0 \begin{pmatrix} \boldsymbol{b}_{m,1} \\ \vdots \\ \boldsymbol{b}_{m,N} \end{pmatrix}, \boldsymbol{T}^1 \begin{pmatrix} \boldsymbol{b}_{m,1} \\ \vdots \\ \boldsymbol{b}_{m,N} \end{pmatrix}, \ldots \boldsymbol{T}^M \begin{pmatrix} \boldsymbol{b}_{m,1} \\ \vdots \\ \boldsymbol{b}_{m,N} \end{pmatrix} \right] \tag{2.20}$$

An analogous conclusion may be drawn for the row vectors in $\boldsymbol{C}$. The dynamics of a controllable and observable system can be characterized by its transfer function. If the system is not totally controllable or observable, then the transfer function characterizes the dynamics of the controllable and observable subspace of a system. [Smi07, p.354] [Gil63]

2.2 Nonlinear Power-Preserving Systems

The designation of nonlinearity is by its very definition pathologically general. This is in stark contrast to the designation of linearity, which is highly specific. The set of nonlinear systems must cover the entire set of conceivable systems, minus the linear subset. It is therefore advantageous to impose additional structural constraints on the system dynamics, if possible, to limit the scope of model term and parameter estimation.

Energy conservation provides a simple, intuitive constraint which arguably already appears in isolated systems. Applying this constraint to the perturbations on the linear model preserves the ability to encode the system's stability or losses in the linear dynamics. Since the stability conditions of the linear model are well understood, this may be preferable to the alternative in some cases. As we will see, the conservation of energy yields a mathematical form which readily lends itself to analysis and interpretation.

Intuitively, the constraint isolates the perturbations on the linear model to interactions between the modes.

2.2.1 Linear Coupling

Consider a model based on simple perturbations of the linear model described in equation (2.3). As a trivial case, those perturbations might be constant. An example of such a system is one that exhibits *linear coupling*. In linear coupling, energy flows between two or more modes, resulting in a constant shift in the complex amplitudes of those modes. The resultant behavior of the system is still linear, and from an analysis perspective, the action of the coupling is indistinguishable from simply moving the location of the poles.

This use of the term "coupling" differs from the use presented in [Smi07, p.368], where the author describes distinct modes with repeated eigenstructure as "coupled." Instead, we propose the term "degenerate" to describe the case where there are repeated eigenvalues or eigenvectors. In the scope of this document, we use the terminology developed e.g. in [Pie74, p.47], where two coupled modes are of different eigenstructure, and energy transfers between them.

The magnitudes and directions of energy transfer due to coupling will be represented as a linear transformation. To describe our constraint formally, let the operation $||\boldsymbol{x}|| \equiv \sqrt{x_1\bar{x}_1, x_2\bar{x}_2, \ldots, x_N\bar{x}_N}$, $\forall x \in \mathbb{C}$, also known as the Euclidean norm of a vector. In order for energy to be conserved in this model, this norm must be preserved, and therefore, in the modal basis, the coupling matrix $\boldsymbol{U}$ must be unitary:

$$\begin{aligned} \boldsymbol{U}\boldsymbol{U}^* &= \boldsymbol{U}^*\boldsymbol{U} = \boldsymbol{I} \\ \boldsymbol{U}^* &= \boldsymbol{U}^{-1} \end{aligned} \tag{2.21}$$

The coupling matrix $\boldsymbol{U}$ acts as a scale-preserving linear transformation on $\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1}$, which is applied prior to the potentially lossy linear transformation $\boldsymbol{\Lambda}$, which together describe the change in state from time $k-1$ to k:

$$\begin{aligned} \overset{\boldsymbol{S}}{\boldsymbol{x}}_k &= \boldsymbol{\Lambda}\boldsymbol{U}_k\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1} + \overset{\boldsymbol{S}}{\boldsymbol{B}}\boldsymbol{u}_k + \overset{\boldsymbol{S}}{\boldsymbol{C}}\boldsymbol{\xi}\boldsymbol{p}_k \\ \boldsymbol{y}_k &= \overset{\boldsymbol{S}}{\boldsymbol{W}}\overset{\boldsymbol{S}}{\boldsymbol{x}}_k + \boldsymbol{\xi}\boldsymbol{m}_k \end{aligned} \tag{2.22}$$

Here $\boldsymbol{S}$ is the eigenvector basis and $\boldsymbol{\Lambda}$ the matrix with $\lambda_1, \lambda_2, \ldots, \lambda_{N\theta}$ on the diagonal. Net energy losses are still entirely encoded in $||\boldsymbol{\Lambda}||$, an effect of $\boldsymbol{U}$'s structural constraint– we have required that $||\boldsymbol{U}|| = \boldsymbol{I}$. In a sense, the coupling matrix $\boldsymbol{U}$ is an adjacency matrix with complex weights, describing a time-invariant, directed network of energy flows amongst the modes.

Recall the constraints, outlined in section 2.1.2.4, that a system realization must satisfy to be controllable and observable. With the addition of $\boldsymbol{U}$, we now have an additional subsystem–albeit a simple one for now–that is not guaranteed to be either. That being said, we may relax this requirement in favor of the constraint that the uncontrollable or unobservable subsystem merely be stable. If this weaker condition is met, the subspace that is controllable or observable will be interpretable under the same analysis techniques as a fully controllable or observable system. We therefore require the same constraints on the parameters $\boldsymbol{W}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ as outlined in section 2.1.2.4 for a system that is controllable and observable. [Gil63]

In the presence of no additional information, the best candidate model for a system such as (2.22) is identical to the linear case. We are unlikely to be able to isolate the effects of $\boldsymbol{U}$ from $\boldsymbol{\Lambda}$, and thus, as stated above, the system is indistinguishable from a linear system with decoupled poles in different locations. The terms in $\boldsymbol{\Lambda}$ as well as $\boldsymbol{S}$ will absorb the terms in $\boldsymbol{U}$. However, if the model parameters have been previously

identified, and the modal basis is known, then any later changes to the coupling of modes may be estimated using this model, simply by plugging in the previously estimated values for $\boldsymbol{S}$ and $\boldsymbol{\Lambda}$ and solving for $\boldsymbol{U}_k$. Note that $\boldsymbol{U}_k$ may now vary in time, and indeed this is the assumption that allows $\boldsymbol{U}$ to be observed at all.[Smi07, p.354]

Notice that the unitary structure of $\boldsymbol{U}$ is only guaranteed in a diagonalized basis. In fact, it is the choice of basis which allows us to interpret $\boldsymbol{U}$ as an adjacency matrix for an energy-conserving directed graph, and this choice will in turn affect the units with which these flows are measured. That $\boldsymbol{U}$ is unitary is equivalent to the constraint that $\overset{\boldsymbol{E}}{\boldsymbol{U}}$ must be orthogonal. Such a constraint would amount to precisely the same result on the system's dynamics, while keeping the weight coefficients in the adjacency matrix real-valued. One might say a unitary matrix is unitarily similar to an orthogonal matrix, and in our case we use the convention that the basis be $\boldsymbol{E}$. In fact, we will use several different bases to observe $\boldsymbol{U}$'s power-preserving constraint and its effects on the system.

2.2.2 Nonlinear Coupling

One might imagine a simple form of nonlinear coupling wherein the inter-modal energy flows depend upon the modal responses themselves. In this case, the same constraint of Euclidean norm-conservation is required, and therefore $\boldsymbol{U}$ will remain unitary. However, now $\boldsymbol{U}_k$ is a function of $\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1}$, the previous modal state vector. To construct such a matrix, we can use the identity

$$\boldsymbol{U}_k \equiv e^{i\boldsymbol{H}_k} \tag{2.23}$$

where the hermitian matrix $\boldsymbol{H}_{ab} = \bar{\boldsymbol{H}}_{ba}$. To map the previous modal state vector $\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1}$ into $\boldsymbol{H}$ requires the following operation:

$$\boldsymbol{H}_k \equiv -i\sum_c (\overset{\boldsymbol{S}}{\mathcal{M}}_c(\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1})_c) \tag{2.24}$$

where $\overset{\boldsymbol{S}}{\mathcal{M}}$ is a 3-tensor whose complex entries describe the effect of each modal response on each energy flow between the modes. Substituting this expansion for $\boldsymbol{H}_k$ in equation (2.23), and placing the resulting unitary matrix $\boldsymbol{U}_k$ into equation (2.22) gives us a state-space form for this new model:

$$\begin{aligned} \overset{\boldsymbol{S}}{\boldsymbol{x}}_k &= \boldsymbol{\Lambda} e^{i\boldsymbol{H}_k} \overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1} + \overset{\boldsymbol{S}}{\boldsymbol{B}}\boldsymbol{u}_k + \overset{\boldsymbol{S}}{\boldsymbol{C}}\boldsymbol{\xi}\boldsymbol{p}_k \\ \boldsymbol{y}_k &= \overset{\boldsymbol{S}}{\boldsymbol{W}} \overset{\boldsymbol{S}}{\boldsymbol{x}}_k + \boldsymbol{\xi}\boldsymbol{m}_k \end{aligned} \tag{2.25}$$

Applying the definition of $\boldsymbol{H}$ shown in equation (2.24), we can expand the above system to show its non-linear dependence on terms in $\boldsymbol{x}$:

$$\overset{\boldsymbol{S}}{\boldsymbol{x}}_k = \boldsymbol{\Lambda}\boldsymbol{S}^{-1} \exp\Big(\sum_c (\mathcal{M}_c(\boldsymbol{S}\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1})_c)\Big)\boldsymbol{S}\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1} + \overset{\boldsymbol{S}}{\boldsymbol{B}}\boldsymbol{u}_k + \overset{\boldsymbol{S}}{\boldsymbol{C}}\boldsymbol{\xi}\boldsymbol{p}_k \tag{2.26}$$

Observing how $\boldsymbol{S}$ propagates through the equations, we can define the change-of-basis on the new parameter $\mathcal{M}$:

$$\overset{\boldsymbol{S}}{\mathcal{M}}_f \equiv \sum_c ((\boldsymbol{S}^{-1}\mathcal{M}_c\boldsymbol{S})\boldsymbol{S}_{cf}) \tag{2.27}$$

Using the simplified notation for similarity transformations first introduced in section

2.1.2, we derive the system equations for the nonlinear model in the modal basis:

$$\begin{aligned}\overset{\boldsymbol{S}}{\boldsymbol{x}}_k &= \boldsymbol{\Lambda}\exp\Big(\sum_c \big(\overset{\boldsymbol{S}}{\mathcal{M}}_c(\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1})_c\big)\Big)\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1} + \overset{\boldsymbol{S}}{\boldsymbol{B}}\boldsymbol{u}_k + \overset{\boldsymbol{S}}{\boldsymbol{C}}\boldsymbol{\xi}\boldsymbol{p}_k \\ \boldsymbol{y}_k &= \overset{\boldsymbol{S}}{\boldsymbol{W}}\overset{\boldsymbol{S}}{\boldsymbol{x}}_k + \boldsymbol{\xi}\boldsymbol{m}_k\end{aligned} \tag{2.28}$$

We will explore the rich form of this model by constructing incrementally more general representations. Since we have designed the 3-tensor $\mathcal{M}$ to capture all the model's deviations from linearity, we will be especially careful to watch its expansion from the perspective of various coordinate systems. To simplify the equations, we will also use a general random variable $\boldsymbol{\xi}_k \equiv \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{C}\boldsymbol{\xi}\boldsymbol{p}_k$ to model the input terms and parameters stochastically, and furthermore assume there is no measurement noise:

$$\begin{aligned}\overset{\boldsymbol{S}}{\boldsymbol{x}}_k &= \boldsymbol{\Lambda}\exp\Big(\sum_c \big(\overset{\boldsymbol{S}}{\mathcal{M}}_c(\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1})_c\big)\Big)\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1} + \overset{\boldsymbol{S}}{\boldsymbol{\xi}}_k \\ \boldsymbol{y}_k &= \overset{\boldsymbol{S}}{\boldsymbol{W}}\overset{\boldsymbol{S}}{\boldsymbol{x}}_k\end{aligned} \tag{2.29}$$

We will add some of these terms back into the equation only once we need them. Eventually, in chapter 3, we will consider methods for estimating its entries, given recordings of the output signals.

2.2.3 The General Structure of $\mathcal{M}$

Given that the model's deviation from linearity is power-preserving, we would like to develop a basis under which the inter-modal transfer of power can be observed. We posit the real modal basis, determined by $\boldsymbol{SE}$:

$$\overset{\boldsymbol{SE}}{\mathcal{M}}_f \equiv \sum_c \big((\boldsymbol{E}^*\boldsymbol{S}^{-1})\mathcal{M}_c(\boldsymbol{SE})(\boldsymbol{SE})_{cf}\big) \tag{2.30}$$

In this basis, the constraint that the nonlinear coupling must be power-preserving results in the following structure in the parameter that encodes it:

$$\overset{\boldsymbol{SE}}{\mathcal{M}}_{abc} = -\overset{\boldsymbol{SE}}{\mathcal{M}}_{bac} \tag{2.31}$$

$$\overset{\boldsymbol{SE}}{\mathcal{M}}_{abc} + \overset{\boldsymbol{SE}}{\mathcal{M}}_{bca} + \overset{\boldsymbol{SE}}{\mathcal{M}}_{cab} = 0 \tag{2.32}$$

which are reminiscent of Kirchoff's circuit laws, or other analogous laws based on resource conservation in a network. The ab sheets of $\mathcal{M}$ are essentially adjacency matrices, which determine the flow of energy from one state variable to another, with b determining the source and a the destination of the flow. At any given time k, the previous state vector $\boldsymbol{x}_{k-1}$ determines the combination of these coupling effects via the c axis. Seen this way, the structure described in (2.31) and (2.32) is roughly equivalent to Kirchoff's Voltage Law in that the 2- and 3-cycles of indices describe loops in the network, which therefore must be equipotentials.

2.2.4 The Complex-Valued Scalar Form

Consider the model's simplest structural configuration: a complex-valued model with a single degree of freedom, where the previous output perturbs the phase of the current output. This view is especially clear since we are permitted to observe the effects of $\boldsymbol{U}$'s structure without the added clutter of the similarity transformations that will reveal it in the wild. The difference equation then becomes:

$$y_k = \lambda \exp\Big((i\mathrm{Re}\,(\mu y_{k-1}))\Big) y_{k-1} + \xi_k \tag{2.33}$$

Note the model terms y_k and ξ_k are now complex scalars, not real vectors, and no longer need to be transformed into the state space at all, thus $y = x$, momentarily. Intuitively,

this has the physical interpretation that we are measuring state directly, which in the one-pole case would mean we are measuring both position and instantaneous velocity, or any linear transformation of the two, or in fact, any combination of a measurable unit and its time-derivative. λ is also a complex scalar, since the system is already effectively diagonalized, by virtue of the fact that there is only one state. Note also how μ becomes a complex scalar instead of a tensor, and that the hermitian matrix $\boldsymbol{H}$ from equation (2.25) simplifies to a real scalar.

We pause for a moment to consider the interpretation of the term μ from a computer-music perspective. By analogy to the treatment in section 2.1.2, as long as we remain in the diagonalized basis, the effect that μ, a complex scalar, has on model dynamics can best be understood in terms of polar coordinates. It describes a signal-dependent phase shift in the feedback path of the one-pole filter. The magnitude, $||\mu||$, describes the phase shift in either direction, measured in radians. The angle, $\angle\mu = \tan^{-1}\frac{\mathrm{Im}(\mu)}{\mathrm{Re}(\mu)}$, determines the phase of the perturbation relative to the phase of the output signal. When driven by white noise, the resulting magnitude spectra produced by models of this form exhibit sidebands at integer multiples of $\omega = \angle\lambda$, including 0, as $||\mu||$ increases.

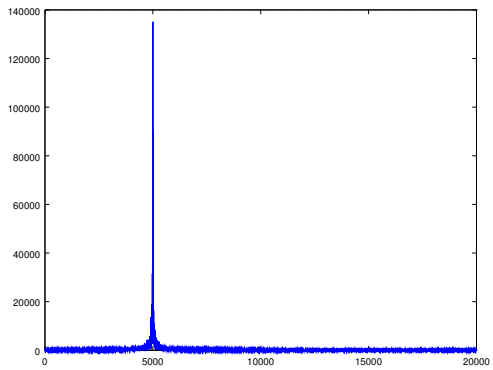

Figure 2.1: Magnitude spectrum of one pole form, $\angle\lambda = \frac{\pi}{2}$, $||\lambda|| = 0.999$, $\angle\mu = 0$, $||\mu|| = 0$. Equivalent to the linear case.

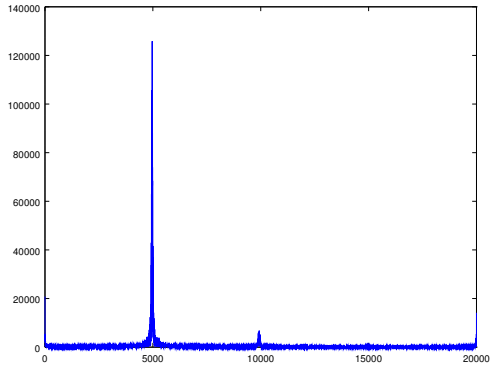

Figure 2.2: Magnitude spectrum of one pole form, $\angle\lambda = \frac{\pi}{2}$, $||\lambda|| = 0.999$, $\angle\mu = 0$, $||\mu|| = 0.005$. Sidebands appear at integer multiples of $\angle\lambda$.

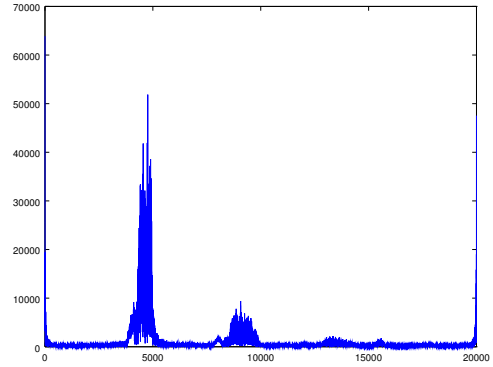

Figure 2.3: Magnitude spectrum of one pole form, $\angle\lambda = \frac{\pi}{2}$, $||\lambda|| = 0.999$, $\angle\mu = 0$, $||\mu|| = 0.015$. Note the sidebands wrap, rather than fold, since the spectrum is single-sided. As the spectrum broadens, the peaks diminish, and power is preserved.

Such a model can be used as a sound synthesis technique, wherein the spectra produced are the result of filter frequency modulation with operator feedback. Framed in this way, this family of techniques has several precedents, all of which share the work of John Chowning as an antecedent.[Cho73] While Chowning's article famously describes audio synthesis via frequency modulation, it does not extend the idea to perturbations of a linear filter. However, many later authors have applied filter frequency modulation to synthesis. A brief and by no means exhaustive list might include the following: [Cha85, p.537] [Puc11] [SSP15] Each of these authors uses a different filter topology as its linear

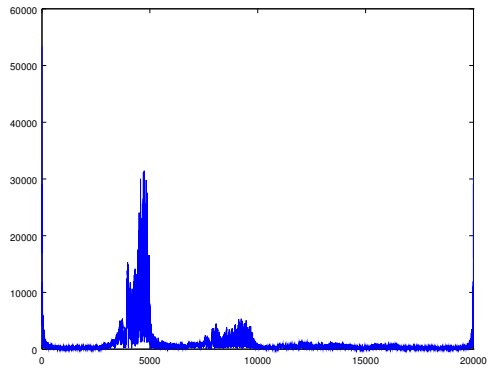

Figure 2.4: Magnitude spectrum of one pole form, $\angle\lambda = \frac{\pi}{2}$, $||\lambda|| = 0.999$, $\angle\mu = 0$, $||\mu|| = 0.02$. As $||\mu||$ increases, energy appears to accumulate at $\omega = 0$ and $\omega = \pi$ radians.

component. Puckette's work, which served as an inspiration for the current endeavor, uses unitary perturbations on a pair of delay lines with feedback. As a special case where the delay lines are of equal length, Puckette briefly discusses the interpretation of such a network as a complex valued, one-pole filter, perturbed by feedback.[Puc11] In Chamberlin's work, the filter is a modified state-variable filter, and the nonlinearity is constrained to waveshaping with an odd function. Both of these authors primarily apply the technique to percussion synthesis, emulating the physics of membranophones or labelophones, whose resonant modes change in response to how they are struck or plucked. Chamberlin, in particular, suggests the use of such a system to partially emulate the sound of a ruler vibrating on a table. Although it is less focused on modeling acoustical sources, Surges et al. mention frequency modulation with feedback as part of a general framework of time-varying second-order all-pass filters, and also recommend the use of unitary matrices to ensure stability. [SSP15]

Despite that the model has constant coefficients, its behavior is not always constant over the same time-scales expected of linear systems of the same order. For certain combinations of coefficients, it is possible for the output signal to transition suddenly into radically different regimes of operation over a period of time. To appreciate this, we

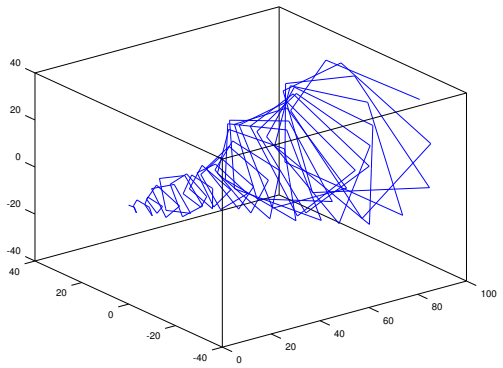

Figure 2.5: Complex time-domain output of one-pole form, $\lambda = i$, $||\mu|| = -0.0175$, $k = 0 \rightarrow 99$

must leave the frequency domain for a moment, and discuss the model's behavior in the time domain with complex-valued dependent variables. The action of the system on an input signal is to rotate around the complex plane at a variable angular velocity centered around $\angle\lambda$. If the conditions allow, the velocity of the rotations may pass through zero, in cases where the nonlinear component of the system is weak, the linear component pushes the angular velocity through zero, and continues behaving like a perturbed linear system, albeit one with only positive frequencies. However, if $||\mu||$ is large enough, the system will stop rotating altogether. This is an unstable equilibrium point in general. In this transitional regime, the model acts as a leaky integrator–a single pole hovering near zero Hz–on the input or stochastic variables. If enough signal accumulates, the model eventually slips into an extremely stable operating regime at a much higher amplitude and frequency. This happens in cases where the model has accumulated a large quantity in a particular direction before folding over.

2.2.5 The Real-Valued Two-Pole Form

To permit real-valued measurements of the system described in equation (2.33), we must supply a second mode, λ_2, such that $\lambda_1 = \bar{\lambda}_2$. The two-pole form of this system

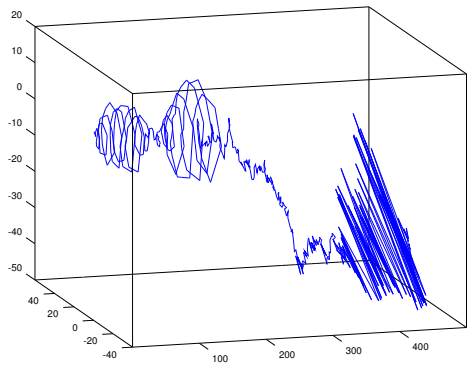

Figure 2.6: Complex time-domain output of one-pole form, $||\lambda|| = 0.9999$, $\angle\lambda = \frac{\pi}{4}$ $||\mu|| = 0.05$, $\angle\mu = -\pi$. $k = 0 \rightarrow 499$

is somewhat close in behavior to the one-pole form. It will differ primarily in the units used in the representation, and the two-sidedness of its spectrum. This affords possible stability improvements for lower values of $\angle\lambda$, with simultaneously higher values of $||\mu||$, as sideband energy no longer simply wraps, but folds as well.

The fact that we have added a row to the state transition equation permits us an opportunity to navigate the formal constraints as we preserve the structure of the model. As our state transition terms remain in modal coordinates, but our observations do not, we begin requiring the use of the modal similarity transformation, and thus $y \neq x$. Furthermore, the state transition coefficient becomes the diagonal matrix $\mathbf{\Lambda}$. The rest of the terms in the state update equation follow suit, increasing their dimensions appropriately:

$$\overset{s}{\begin{bmatrix} x_{1,k} \\ x_{2,k} \end{bmatrix}} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_1 \end{bmatrix} \exp\left(i \overset{s}{\left[\begin{array}{cc} \mu_{1,1,1} & \mu_{1,2,1} \\ \mu_{2,1,1} & \mu_{2,2,1} \\ \hline \mu_{1,1,2} & \mu_{1,2,2} \\ \mu_{2,1,2} & \mu_{2,2,2} \end{array}\right]_{abc}} \overset{s}{\begin{bmatrix} x_{1,k-1} \\ x_{2,k-1} \end{bmatrix}}_c \right) \overset{s}{\begin{bmatrix} x_{1,k-1} \\ x_{2,k-1} \end{bmatrix}} + \overset{s}{\begin{bmatrix} \xi_{1,k} \\ \xi_{2,k} \end{bmatrix}} \tag{2.34}$$

Assuming the diagonal coefficients in $\mathbf{\Lambda}$ do not lie on the real axis, the symmetries of this system are: $\overset{\boldsymbol{S}}{x}_{1,k} = \overset{\bar{\boldsymbol{S}}}{x}_{2,k}$, $\lambda_1 = \bar{\lambda}_2$, $\overset{\boldsymbol{S}}{\xi}_1 = \overset{\bar{\boldsymbol{S}}}{\xi}_2$, and $\overset{\boldsymbol{S}}{\mathcal{M}}_{abc} = \overset{\bar{\boldsymbol{S}}}{\mathcal{M}}_{bac}$. The components in $\overset{\boldsymbol{S}}{\mu}$ once again describe complex weights in an energy-transfer adjacency matrix for a network of complex modes.

The 3-tensor $\mathcal{M}$ is unitarily similar to a sparser, antisymmetric 3-tensor, which we can observe by transforming into the real-valued block-diagonalized basis via the matrix $\boldsymbol{E}$, which we introduced in section 2.1.2.3. In 2-tensors such as $\boldsymbol{R}$, the blocks are 2×2, in 3-tensors such as $\mathcal{M}$, the blocks are $2 \times 2 \times 2$. This change of coordinates gives equation (2.34) the following form:

$$\begin{bmatrix} \boldsymbol{SE} \\ x_{1,k} \\ x_{2,k} \end{bmatrix} = \begin{bmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{bmatrix} \exp \left(\begin{bmatrix} \boldsymbol{SE} \\ 0 & \mu_{1,2,1} \\ \mu_{2,1,1} & 0 \\ \hline 0 & \mu_{1,2,2} \\ \mu_{2,1,2} & 0 \end{bmatrix}_{abc} \begin{bmatrix} x_{1,k-1} \\ x_{2,k-1} \end{bmatrix}_c \right) \begin{bmatrix} \boldsymbol{SE} \\ x_{1,k-1} \\ x_{2,k-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{SE} \\ \xi_{1,k} \\ \xi_{2,k} \end{bmatrix} \quad (2.35)$$

The fact that $\mathcal{M}$ is now sparse and must be anti-symmetric reflects the number of free parameter variables in the system given by our constraints set up in section 2.2. Viewed as an unstructured $N \times N \times N$ 3-dimensional array, $\mathcal{M}$ is quite an expensive term in general– its number of coefficients grows as N^3– but the unitary constraint on $\mathcal{M}$ provides a more parsimonious model than the general case. Observing its actions in this sparse representation exposes their underlying simplicity.

Again, we pause to examine the frequency-domain features of this system. To generate figures 2.7 and 2.8, we set $y_k = \boldsymbol{w}\boldsymbol{S}\boldsymbol{E}\overset{\boldsymbol{SE}}{\boldsymbol{x}}_k$, and that $\boldsymbol{w} = \frac{1}{\sqrt{2}}(1,1)$, yielding scalar output. The sidebands appear at positive and negative integer multiples of $\angle\lambda_1$ and $\angle\lambda_2$. As $||\lambda_1||$ and $||\lambda_2|| \to 1$, the series of sidebands formed by such a system becomes reminiscent of those formed by traditional FM synthesis.

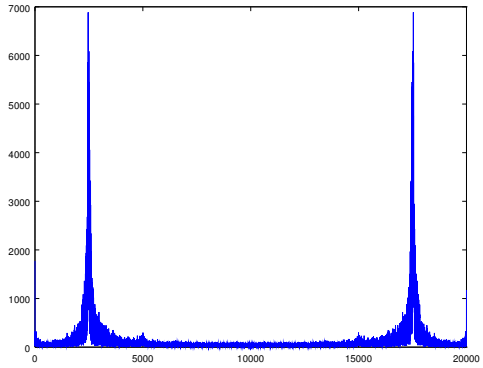

Figure 2.7: Magnitude spectrum of two pole form, $\angle\lambda = \frac{\pi}{2}$, $||\lambda|| = 0.99$, $\mu_{1,2,1} = \mu_{1,2,2} = 0.006$. Note the two sided spectrum.

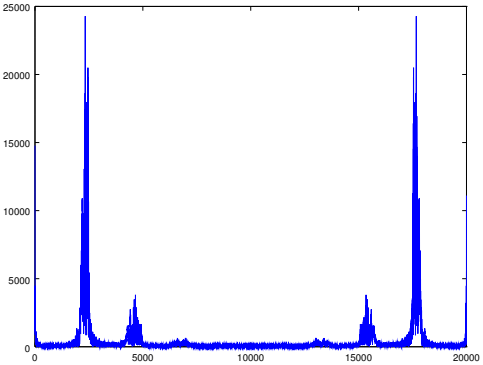

Figure 2.8: Magnitude spectrum of two pole form, $\angle\lambda = \frac{\pi}{2}$, $||\lambda|| = 0.9998$, $\mu_{1,2,1} = \mu_{1,2,2} = 0.006$. Sidebands form without directly increasing $||\mu||$.

2.2.6 The Real-Valued Form with $N > 2$

We continue to extend the model by adding diagonal state elements to $\mathbf{\Lambda}$. To avoid adding coefficients that lie on the real axis, we must add them in conjugate pairs.

For example, when $N = 4$:

$$\overset{\boldsymbol{SE}}{\begin{bmatrix} x_{1,k} \\ x_{2,k} \\ x_{3,k} \\ x_{4,k} \end{bmatrix}} = \begin{bmatrix} \boldsymbol{R}_1 & & \\ & \boldsymbol{R}_2 & \\ & & \end{bmatrix} \exp \left(\begin{bmatrix} 0 & \mu_{1,2,1} & \mu_{1,3,1} & \mu_{1,4,1} \\ \mu_{2,1,1} & 0 & \mu_{2,3,1} & \mu_{2,4,1} \\ \mu_{3,1,1} & \mu_{3,2,1} & 0 & \mu_{3,4,1} \\ \mu_{4,1,1} & \mu_{4,2,1} & \mu_{4,3,1} & 0 \\ \hline 0 & \mu_{1,2,2} & \mu_{1,3,2} & \mu_{1,4,2} \\ \mu_{2,1,2} & 0 & \mu_{2,3,2} & \mu_{2,4,2} \\ \mu_{3,1,2} & \mu_{3,2,2} & 0 & \mu_{3,4,2} \\ \mu_{4,1,2} & \mu_{4,2,2} & \mu_{4,3,2} & 0 \\ \hline 0 & \mu_{1,2,3} & \mu_{1,3,3} & \mu_{1,4,3} \\ \mu_{2,1,3} & 0 & \mu_{2,3,3} & \mu_{2,4,3} \\ \mu_{3,1,3} & \mu_{3,2,3} & 0 & \mu_{3,4,3} \\ \mu_{4,1,3} & \mu_{4,2,3} & \mu_{4,3,3} & 0 \\ \hline 0 & \mu_{1,2,4} & \mu_{1,3,4} & \mu_{1,4,2} \\ \mu_{2,1,4} & 0 & \mu_{2,3,4} & \mu_{2,4,4} \\ \mu_{3,1,4} & \mu_{3,2,4} & 0 & \mu_{3,4,2} \\ \mu_{4,1,4} & \mu_{4,2,4} & \mu_{4,3,4} & 0 \end{bmatrix}_{abc} \overset{\boldsymbol{SE}}{\begin{bmatrix} x_{1,k-1} \\ x_{2,k-1} \\ x_{3,k-1} \\ x_{4,k-1} \end{bmatrix}}_c \right) \overset{\boldsymbol{SE}}{\begin{bmatrix} x_{1,k-1} \\ x_{2,k-1} \\ x_{3,k-1} \\ x_{4,k-1} \end{bmatrix}} + \overset{\boldsymbol{SE}}{\begin{bmatrix} \xi_{1,k} \\ \xi_{2,k} \\ \xi_{3,k} \\ \xi_{4,k} \end{bmatrix}} \tag{2.36}$$

As we increase the number of states, the term $\overset{\boldsymbol{SE}}{\mathcal{M}}$ gains dense, 2×2 blocks on its off-diagonal positions. In this coordinate system, these blocks are of the form of $\boldsymbol{N} = \begin{pmatrix} n_1 & n_2 \\ -n_2 & n_1 \end{pmatrix}$, but the tensor $\overset{\boldsymbol{SE}}{\mathcal{M}}$ loses its large-scale antisymmetry in *ab*-blocks along the *c* axis. However, the antisymmetric structure of the diagonal blocks, which we will call $\overset{\boldsymbol{SE}}{\boldsymbol{M}}$, are left unchanged from the two-state case.

The $\boldsymbol{N}$ blocks enter $\mathcal{M}$ as follows:

$$\overset{\boldsymbol{SE}}{\begin{bmatrix} x_{1,k} \\ x_{2,k} \\ \vdots \\ x_{N,k} \end{bmatrix}} = \boldsymbol{R}\exp\left(\overset{\boldsymbol{SE}}{\left[\begin{array}{cccc} \boldsymbol{M}_{1,1,1} & \boldsymbol{N}_{1,2,1} & \cdots & \boldsymbol{N}_{1,N,1} \\ \boldsymbol{N}_{2,1,1} & \boldsymbol{M}_{2,2,1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{N}_{N-1,N,1} \\ \boldsymbol{N}_{N,1,1} & \cdots & \boldsymbol{N}_{N,N-1,1} & \boldsymbol{M}_{N,N,1} \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline \boldsymbol{M}_{1,1,N} & \boldsymbol{N}_{1,2,N} & \cdots & \boldsymbol{N}_{1,N,N} \\ \boldsymbol{N}_{2,1,N} & \boldsymbol{M}_{2,2,N} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{N}_{N-1,N,N} \\ \boldsymbol{N}_{N,1,N} & \cdots & \boldsymbol{N}_{N,N-1,N} & \boldsymbol{M}_{N,N,N} \end{array}\right]_{abc}} \overset{\boldsymbol{SE}}{\begin{bmatrix} x_{1,k-1} \\ x_{2,k-1} \\ \vdots \\ x_{N,k-1} \end{bmatrix}}_{c} \right) \overset{\boldsymbol{SE}}{\begin{bmatrix} x_{1,k-1} \\ x_{2,k-1} \\ \vdots \\ x_{N,k-1} \end{bmatrix}} + \overset{\boldsymbol{SE}}{\begin{bmatrix} \xi_{1,k} \\ \xi_{2,k} \\ \vdots \\ \xi_{N,k} \end{bmatrix}} \tag{2.37}$$

While conjugate state variable pairs transfer energy among each other in the $\overset{\boldsymbol{SE}}{\boldsymbol{M}}$ rotation blocks, they remain within a single real mode. Pairs of off-diagonal $2 \times 2 \times 2$ blocks $\overset{\boldsymbol{SE}}{\boldsymbol{N}}_{abc}$ and $\overset{\boldsymbol{SE}}{\boldsymbol{N}}_{bac}$ describe the energy transfer between real modes. The pair of state variables used in one real mode may be scaled differently from those used in another real mode. This loss of absolute scale is a result of several parameters, which encode how the measured modal responses differ from unity. We had implicitly assumed unity scaling with the normalization of the eigenvector basis $\boldsymbol{S}$, described in section 2.13, and required for our modal coordinate system to be unique up to a linear transformation of eigenvectors.

The differences in scale between state variables that correspond to different real modes are encoded by $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{W}$, which we have largely ignored up until this point in the derivation, in the interest of clarity. Recall that $\boldsymbol{B}$ and $\boldsymbol{C}$ are the model parameters

that couple the input term $\boldsymbol{u}_k$, and the process noise term $\boldsymbol{\xi}_k$ with the modes, respectively. Finally, the model parameter $\boldsymbol{W}$, transforms the state of the model into our observations of it.

It is worthwhile to pause once again and consider this system from the perspective of audio synthesis. In the complex modal basis discussed in section 2.2.4, the argument of μ, then a complex scalar, offset the phase of the carrier signal in the single operator frequency modulation with feedback. Now, that very same synthesis parameter is encoded as the angle between the non-zero coefficients $\angle \boldsymbol{M}_{nnc} = \tan^{-1} \frac{\boldsymbol{M}_{nnc}}{\boldsymbol{M}_{nnc+1}}$

Chapter 3

Parameter Estimation on the Model (Analysis)

In this chapter, we derive a method for the estimation of the parameters of the nonlinear model described in chapter 2. As described in section 2.2.2, we would like to estimate the model parameters given only sensor data (possibly multi-channel) over time. We will model the excitation signal as an independent and identically distributed stochastic variable, ξ.

We begin by assuming a linear relationship between the parameters and the terms. This assumption is common in signal processing and system identification, and furthermore, economical to apply and validate in practice. Note that this is not equivalent to assuming that the system itself is linear. We may not require the output vector $\boldsymbol{y}$ to be linear in terms of its previous inputs, outputs, and noise measurements, but only require $\boldsymbol{y}$ to be linear in terms of the generalized parameter vector, θ. [Bil13, p.64] This assumption can be notated as:

$$y = \boldsymbol{p}\theta \tag{3.1}$$

wherein we collect the terms– which are presumed known or fixed at this stage of the

analysis– into $\boldsymbol{p}$, and parameters– which are unknown– into θ. Notice the similarity of equation (3.1) with equation (2.8). We will reintroduce the residual term, $\boldsymbol{e}_{\text{lin}}$, in section 3.1.1, when we will also perform an in-depth analysis of its interpretation and decomposition.

3.1 Solving for the Linear Parameters

Since our synthesis model was presented as a perturbed linear system, we assume that there exist terms in $\boldsymbol{p}$ which are themselves linear with respect to y. We are, after all, looking for the simplest possible explanatory model, and a linear relationship might offer a reasonable first-order approximation to the system. Furthermore, the estimation of linear terms is a well-studied problem, and many robust and economical solutions already exist. [Alp14, p.80]

The problem of parameter estimation, when applied to the model described by the equation (3.1) has infinitely many solutions. This is because there are ostensibly many unknowns (the parameters θ), and only one equation. Such a system is said to be *under-determined*. Therefore, rather than solving the system directly, we must find an optimal set of synthesis parameters that best describes the measurements. To determine how well a particular choice in parameters approximates those of the measured system, we must define an objective function that we either minimize (such as a cost or error function), or maximize (such as an expectation or likelihood function).

We attempt to improve the estimation problem from the indeterminate scalar-valued form in equation (3.1) by collecting observations through time. Formally, this means increasing the number of equations M in the system, until $M \geq N$, and it becomes *over-determined*. The additional rows describe the difference equation with $\boldsymbol{y}_k = \left(y_{k-0}, y_{k-1}, \ldots, y_{k-M}\right)^{\mathrm{T}}$ as the left-hand side. The model term vector $\boldsymbol{p}_k$, to ac-

commodate this increase in dimension, must become the matrix $\boldsymbol{P}$, whose row vectors $\boldsymbol{p}_m = \boldsymbol{p}_{K-(m-1)}$. The matrix form of the difference equation (2.8) can thus be written as:

$$\boldsymbol{y}_k = \boldsymbol{P}\theta + \boldsymbol{e}_{\text{lin}k}. \tag{3.2}$$

In the case where $M = N$, although the dimensions seem ideal, and a solution is algebraically defined, there is a good chance that there will be no solution. This could be because of noise in the measurements or in the model terms, a model term we haven't considered, or any other inadequacy of our model, captured in this case by the column vector $\boldsymbol{e}_k$. In this case, we use the residual term e_{lin} to denote the component of the output not explained by the linear component of the model. So, we have gone from a case where we have infinitely many solutions to no solutions. Clearly, there is more work to do.

3.1.1 The Least Squares Residual

The method of *Least Squares* was first developed by Karl Friedrich Gauss, circa 1809. [Nor86, p.60] The goal of ordinary least squares is to find an optimal solution of the form $\hat{\theta} = \left(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_N\right)^T$ to the system of equations $\boldsymbol{P}\theta = \boldsymbol{y}$, where no direct solution exists. [Str09, p.218] The residual $\boldsymbol{e}$ is a column vector whose coefficients contain the projection of θ into the nullspace of $\boldsymbol{P}^{\text{T}}$:

$$\begin{pmatrix} e_{\text{lin}K-0} \\ e_{\text{lin}k-1} \\ \vdots \\ e_{\text{lin}k-(M-1)} \end{pmatrix} = \begin{pmatrix} y_{k-0} \\ y_{k-1} \\ \vdots \\ y_{k-(M-1)} \end{pmatrix} - \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,N} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M,1} & p_{M,2} & \cdots & p_{M,N} \end{pmatrix} \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_{N\theta} \end{pmatrix}$$

$$\boldsymbol{e}_{\text{lin}k} = \boldsymbol{y}_k - \boldsymbol{P}\hat{\theta} = \boldsymbol{y}_k - \hat{\boldsymbol{y}}_k \tag{3.3}$$

where $\hat{\boldsymbol{y}}_k = \boldsymbol{P}\hat{\theta}$ is the estimated value of $\boldsymbol{y}_k$, given $\hat{\theta}$. For just a moment, we assume $\boldsymbol{P}$

is square, so that all operations are defined.

3.1.2 Ordinary Least Squares

The column vector $\boldsymbol{e}_{\text{lin}k}$ provides a convenient distance measure between the estimated and actual values of $\boldsymbol{y}_k$. Assuming $\boldsymbol{y}_k$ is centered at zero, the total sum of squares of the residual approximates the amount of measured variance not explained by the model.

$$E\left(\boldsymbol{e}_{\text{lin}}{}^T\boldsymbol{e}_{\text{lin}}\right) = E\left(\sum_{m=0}^{M} e_{\text{lin}m}^2\right) = E\left(||\boldsymbol{y} - \boldsymbol{P}\hat{\theta}||^2\right) \tag{3.4}$$

The *mean squared error* (MSE), $E\left(\boldsymbol{e}^T\boldsymbol{e}\right)$ is the error measure we choose to minimize, both because of its extremely tractable and interpretable form, and also because it penalizes large errors at a higher rate than small errors. Note that this second criterion may also be a liability if there are large outliers in the dataset. [Nor86, p.61] Nonetheless, via equation (3.4) our optimization problem can now be formally specified as:

$$\hat{\theta} = \arg\min_{\theta} E\left(||\boldsymbol{y} - \boldsymbol{P}\hat{\theta}||^2\right) \tag{3.5}$$

We proceed by taking the gradient of the MSE with respect to θ. This results in the familiar closed-form solution, otherwise known as the *normal equations*:

$$\hat{\theta} = \left(\boldsymbol{P}^T\boldsymbol{P}\right)^{-1}\boldsymbol{P}^T\boldsymbol{y} \tag{3.6}$$

Note that, assuming the distribution of $\boldsymbol{P}$ is mean-zero, this form describes the unknowns in relation to the approximated *auto-covariance*, $\boldsymbol{P}^T\boldsymbol{P}$, and the approximated *cross-covariance*, $\boldsymbol{P}^T\boldsymbol{y}$, as $M \to \infty$.

The technique of least squares can be readily applied to the parameter estimation of linear systems by directly estimating the non-trivial elements in the linear state

transition matrix $\boldsymbol{T}$. Because we have only collected terms concerning previously observed outputs $\boldsymbol{y}$, this matrix $\boldsymbol{T}$ is similar to one whose non-trivial elements consist of the negated coefficients of the polynomial $A(z)$. Furthermore, if the input term matrix $\boldsymbol{U}_k$ has been collected, an analogous process may be used to derive a matrix similar to $\boldsymbol{WB}$, assuming the conditions described in equations (2.20) and (2.19) have been met.

In the present usage, the general term matrix $\boldsymbol{P}$ might be substituted with :

$$\boldsymbol{Y}_k \equiv \begin{bmatrix} y_{1,k-M-1} & y_{2,k-M-1} & \cdots & y_{N,k-M-1} \\ y_{1,k-M} & y_{2,k-M} & \cdots & y_{N,k-M} \\ \vdots & \vdots & & \vdots \\ y_{1,k} & y_{2,k} & \cdots & y_{N,k} \end{bmatrix} \tag{3.7}$$

where the columns of $\boldsymbol{Y}$ might contain output measurements which have been lagged in time, and if they are available, distributed in space. We will discuss the effects of generalizing the full algorithm to *multiple input, multiple output (MIMO)* systems a bit later in section 3.1.3, and will explore the impact of increasing system dimensionality on predictive acuity in chapter 4. By equating the term matrix with previous measurements, we use ordinary least squares to find a estimated autoregressive parameter matrix, which in our previous construction in chapter 2 we had called the state transition matrix $\boldsymbol{T}$. This estimated state transition matrix is similar to a matrix with the following general construction:

$$\hat{\boldsymbol{T}} = (\boldsymbol{Y}_{k-1}^{\mathrm{T}} \boldsymbol{Y}_{k-1})^{-1} \boldsymbol{Y}_{k-1}^{\mathrm{T}} \boldsymbol{Y}_k \tag{3.8}$$

and is the basis for the technique known as *linear predictive coding (LPC)*. Note that the columns of $\boldsymbol{Y}_k$ must be mean-centered. If the columns contain DC information, LPC will express this by placing a pair of poles on the real axis. This essentially wastes two poles of the analysis, and if there are insufficient remaining poles to describe the AC behavior

of the signal, the accuracy of these remaining poles will be compromised. Thus, in order for equation (3.8) to give unbiased estimates, we remove the DC component from the columns of $\boldsymbol{Y}_k$. However, in later parts of the analysis, particularly looking ahead to section 3.3, we will need to add the bias term back.

In canonical audio DSP literature, this technique is typically applied to describe SISO systems, however the inclusion of spatially distributed measurements is germane to other applications, and, as we shall observe, has much to offer the audio use case as well. If we similarly populate the output term matrix $\boldsymbol{Y}_k$ with lagged and spatially distributed output measurements, the result of equation (3.8) will be numerically very close to:

$$\hat{\boldsymbol{T}} \sim \left(\begin{array}{c|c} 0 & I \\ \hline \multicolumn{2}{c}{\hat{\boldsymbol{\Theta}}} \end{array} \right) \tag{3.9}$$

with very small additional variance. The identity block, offset by a block of zeros, performs the time-shifting operation of copying the previous measurements (of dimension M) a single time-step into the past. The prediction block at the bottom, comprised of the coefficients of output polynomial $\boldsymbol{A}$, predicts the same number of measurements that the time-shifting block discards. The literature occasionally absorbs the negative sign present from the difference equation (2.4) into the autoregressive parameter coefficients, whereas we simply leave the negative sign in the difference equation for clarity.

The picture in equation (3.9) is only approximate. The coefficients which ought to contain zeros, in the time-shifting block for example, generally contain very small values instead. In practice, it might be prudent to send these very small values to zero, as they may negatively impact numerical precision. The deviation from sparseness can be viewed as average corrections (over the entire analysis window) which the system must apply in order to adequately predict earlier time-steps from their predecessors. This can

be a sign of nonlinearity or of cross-covariances in $\boldsymbol{\xi}$. In the linear case, these deviations are due solely to the expected value of $\boldsymbol{\xi}$, which we specified should approach zero as $K \to \infty$. To ignore the near-zero terms is to assume that the length of the analysis window is sufficiently large for the product $\boldsymbol{Y}_{k-1}^{\mathrm{T}} \boldsymbol{Y}_{k-1}$ to approximate the autocovariance matrix of $\boldsymbol{Y}_{k-1}$. If we only needed to predict the most recent output of the system given the most recent information, we might ignore these corrections. In the general case, however, we would like to know how the current estimate of the system relates to prior measurements, and this deviation in $\boldsymbol{\xi}$ provides us with important information.

We are now in a position to offer a view of the linear residual $\boldsymbol{e}_{\mathrm{lin}}$ as a function of the measured terms, and its relationship to the higher order terms:

$$\boldsymbol{Y}_k = \hat{\boldsymbol{T}} \boldsymbol{Y}_{k-1} + \boldsymbol{e}_{\mathrm{lin}k} \tag{3.10}$$

$$\boldsymbol{e}_{\mathrm{lin}k} = h.o.t. + \boldsymbol{\xi}_k \tag{3.11}$$

We will subsequently decompose this linear residual term further to extract any higher-order terms that it might contain. To accomplish this, we must examine $\hat{\boldsymbol{T}}$. In the case where we analyze the measurements of M microphones, looking for N linear parameters,

$\hat{\boldsymbol{T}}$ has the following structure:

$$\hat{\boldsymbol{T}} = \underbrace{\left(\begin{array}{ccccccc} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & & 0 & 0 & 1 & \ddots & \vdots \\ \vdots & & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \dots & 0 & 0 & 1 \end{array}\right)}_{M} \Bigg\} N-M$$

$$\underbrace{\left(\begin{array}{ccccc} \hat{a}_{11} & \hat{a}_{12} & \hat{a}_{13} & \dots & \hat{a}_{1N} \\ \vdots & \vdots & \vdots & & \vdots \\ \hat{a}_{M1} & \hat{a}_{M2} & \hat{a}_{M3} & \dots & \hat{a}_{MN} \end{array}\right)}_{N} \Bigg\} M \tag{3.12}$$

Notice that the time-shifting block will disappear entirely if $M = N$. This condition is satisfied when the number of linear features matches the number of microphones being used to measure the system, which in general may be quite rare. This is because a real oscillating system exhibits modal behavior in families of modes, which themselves are arrayed in infinite series. However, the dimensionality of the problem– and thus, the time and space complexity of the algorithm– grows more quickly with N than it does with K. In general, the higher frequency modes become progressively more damped as their frequency increases, resulting in the familiar $\frac{1}{f}$ curve. As a result, these extremely well-damped modes may have a time constant far too small to measurably impact the dynamics of the system. Furthermore, since time is a dimension over which error can accumulate, unlike space, some very interesting properties are theoretically possible when the number of measurements per time-step - M - increases.

3.1.3 State-Space Representation of OLS Estimates

From the estimated state transition matrix $\hat{\boldsymbol{T}}$, we can derive approximations of $\boldsymbol{\Lambda}$, $\boldsymbol{S}$, $\boldsymbol{R}$, and $\overset{\boldsymbol{SE}}{\boldsymbol{x}}$:

$$\hat{\boldsymbol{\Lambda}} = \hat{\boldsymbol{S}}^{-1}\hat{\boldsymbol{T}}\hat{\boldsymbol{S}} \tag{3.13}$$

$$\hat{\boldsymbol{R}} = \boldsymbol{E}^{*}\hat{\boldsymbol{\Lambda}}\boldsymbol{E} \tag{3.14}$$

$$\overset{\boldsymbol{SE}}{\boldsymbol{x}} = \boldsymbol{E}^{*}\boldsymbol{S}^{-1}\boldsymbol{y} \tag{3.15}$$

By diagonalizing the linear state variables, we find a basis for analyzing the linear contributions to the system behavior independently. We opt for the real-valued, block-diagonal representation which transforms the eigenvalue matrix $\hat{\boldsymbol{\Lambda}}$ into the rotation matrix $\hat{\boldsymbol{R}}$ using the same conventions we specified in section 2.1.2.3: our rotations are active, in the clockwise direction, applied by pre-multiplication with the state vector, and our coordinate system is right-handed.

We currently ignore the effect of $\boldsymbol{W}$ on our measurements. Our initial assumption that an autoregressive analysis of our measurements at the system's outputs can result in an unbiased estimate of the state vector disregards the possibility that the independent components of the linear state may have been recorded at different levels. In general this is a shortcoming of LPC, which functionally fits an all-pole filter cascade to the system measurements. This is because in a purely autoregressive model, these variations in cross-modal power level may be interpreted as variations of filter damping, if the variations are large enough to disrupt the spectral envelope. [RP92, p.310] Without adding zeros to the transfer function, this is the best we can do, if there is a component of the system output measurements that relies on time lags, rather than spatial distribution. The sensitivity to variations of this kind increases considerably if $\boldsymbol{W}$ is a row-vector. In this case, a recursive technique such as *Prony's method* may be employed to obtain a

complete estimate of the transfer function, including zeros. [p.393]smith2007

Interestingly enough, theoretically this limitation could also be mitigated for nonlinear systems if we employ the same number of microphones as there are linear state variables. Under this condition, the expression of power variations is entirely encoded in the magnitude of the column vectors in the eigenbasis. In the purely linear case, where the state variable estimates are scale-invariant, we are unable to retrieve this latent information. However, in a complete pass of the algorithm we currently present, we retrieve these scale factors in the process of solving for $\hat{\mathcal{M}}$. In fact, because of the structural constraints we apply to $\hat{\mathcal{M}}$, the search for these scale factors becomes essential in assuring the stability of the estimates. We call this scaling basis $\boldsymbol{V}$, and a technique for estimating it is presented in section 3.3.4.

3.2 Linearization of the Parameters

Since our model was designed in section 2.2.2 as a net-power-preserving transformation, the mean-squared error is an appropriate function to minimize during parameter estimation. If the model were linear with respect to the parameters, we could exploit the ordinary least-squares formulation to derive some information about the system, if not solve it outright. However, it is evident from equations (2.24) and (2.33) that our model is not linear in the parameters. Because of this, the initial task will consist of deriving a sufficiently accurate approximation to the model which satisfies this constraint. Once the model is linearized with respect to θ, we begin fitting estimates for parameters, starting from first-order estimates and subsequently turning our attention to higher-order terms.

3.2.1 A 1st-Order Nonlinear Approximation

We linearize the exponential term in our model about $x = 0$ with a first-order Taylor series approximation. The error is:

$$1 + x - e^x = \frac{x^2}{2} + O\left(x^3\right) \tag{3.16}$$

This suggests that such an approximation may yield a reasonable linearization with respect to the parameters provided $|\boldsymbol{x}_k|$ is small. The resulting estimate for the unitary coupling term $\boldsymbol{U}$ is:

$$\boldsymbol{U}_k \approx \boldsymbol{I} + \sum_c (\overset{\boldsymbol{S}}{\mathcal{M}}_c (\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1})_c) \tag{3.17}$$

Applying the above approximation to the system described in equation (2.29) results in the following linear-in-the-parameters system:

$$\begin{aligned} \overset{\boldsymbol{S}}{\boldsymbol{x}}_k &\approx \boldsymbol{\Lambda}(\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1} + \sum_{bc} (\overset{\boldsymbol{S}}{\mathcal{M}}_{bc} (\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1})_c (\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1})_b)) + \overset{\boldsymbol{S}}{\boldsymbol{\xi}}_k \\ \boldsymbol{y}_k &= \overset{\boldsymbol{S}}{\boldsymbol{W}} \overset{\boldsymbol{S}}{\boldsymbol{x}}_k \end{aligned} \tag{3.18}$$

At this point, we can define a complex-valued basis for our first-order nonlinear approximation of the system, under which the state transition matrix $\boldsymbol{T}$ is diagonalized. We observe explicitly how the similarity transformation encoded by $\boldsymbol{S}$ is applied to each object of interest in the system:

$$\overset{\boldsymbol{S}}{\boldsymbol{x}}_k \approx \boldsymbol{S}^{-1}\boldsymbol{T}\boldsymbol{S}\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1} + \sum_{abc} ((\boldsymbol{S}^{-1}\boldsymbol{T})_a \mathcal{M}_{abc} (\boldsymbol{S}\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1})_c (\boldsymbol{S}\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1})_b) + \boldsymbol{S}^{-1}\boldsymbol{\xi}_k \tag{3.19}$$

Once again, applying the simplification in terms introduced in equation (2.11) and again in equation (2.27) results in the following:

$$\overset{\boldsymbol{S}}{\boldsymbol{x}}_k \approx \boldsymbol{\Lambda}\boldsymbol{S}\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1} + \sum_{abc}((\boldsymbol{\Lambda}\boldsymbol{S}^{-1})_a \mathcal{M}_{abc}(\boldsymbol{S}\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1})_c(\boldsymbol{S}\overset{\boldsymbol{S}}{\boldsymbol{x}}_{k-1})_b) + \overset{\boldsymbol{S}}{\boldsymbol{\xi}}_k \tag{3.20}$$

Which shows us how $\mathcal{M}$ transforms under the change-of-basis defined by $\boldsymbol{S}$ to form $\overset{\boldsymbol{S}}{\mathcal{M}}$. Note the convenient similarity to the exponential form described in equation (2.27). From this representation in equation (3.20) above, we can derive the block-diagonal form:

$$\overset{\boldsymbol{SE}}{\boldsymbol{x}}_k \approx \boldsymbol{E}^*\boldsymbol{S}^{-1}\boldsymbol{T}\boldsymbol{S}\boldsymbol{E}\overset{\boldsymbol{SE}}{\boldsymbol{x}}_{k-1} + \sum_{abc}(\boldsymbol{E}^*\boldsymbol{S}^{-1}\boldsymbol{T}_a\mathcal{M}_{abc}(\boldsymbol{S}\boldsymbol{E}\overset{\boldsymbol{SE}}{\boldsymbol{x}}_{k-1})_c(\boldsymbol{S}\boldsymbol{E}\overset{\boldsymbol{SE}}{\boldsymbol{x}}_{k-1})_b) + \overset{\boldsymbol{SE}}{\boldsymbol{\xi}}_k$$

The transformation rules remain salient for this change-of-basis as well. Consolidating the transformations on the state transition matrix and defining $\boldsymbol{R}$ as in equation (2.16) yields:

$$\overset{\boldsymbol{SE}}{\boldsymbol{x}}_k \approx \boldsymbol{R}\overset{\boldsymbol{SE}}{\boldsymbol{x}}_{k-1} + \sum_{abc}(\boldsymbol{R}(\boldsymbol{E}^*\boldsymbol{S}^{-1})_a\mathcal{M}_{abc}(\boldsymbol{S}\boldsymbol{E}\overset{\boldsymbol{SE}}{\boldsymbol{x}}_{k-1})_c(\boldsymbol{S}\boldsymbol{E}\overset{\boldsymbol{SE}}{\boldsymbol{x}}_{k-1})_b) + \overset{\boldsymbol{SE}}{\boldsymbol{\xi}}_k \tag{3.21}$$

which is the form that permits us to view the effects of the similarity transformation into the block-diagonal real-valued basis on $\mathcal{M}$ in our first-order nonlinear approximation:

$$\overset{\boldsymbol{SE}}{\boldsymbol{x}}_k \approx \boldsymbol{R}\overset{\boldsymbol{SE}}{\boldsymbol{x}}_{k-1} + \sum_{abc}(\boldsymbol{R}_a\overset{\boldsymbol{SE}}{\mathcal{M}}_{abc}(\overset{\boldsymbol{SE}}{\boldsymbol{x}}_{k-1})_c(\overset{\boldsymbol{SE}}{\boldsymbol{x}}_{k-1})_b) + \boldsymbol{E}^*\boldsymbol{S}^{-1}\overset{\boldsymbol{SE}}{\boldsymbol{\xi}}_k \tag{3.22}$$

We can now see that this approximate system is still nonlinear in the state variables $\boldsymbol{x}$ and in the linear rotation matrix $\boldsymbol{R}$. However, it is now linear in the nonlinear coupling tensor $\overset{\boldsymbol{SE}}{\mathcal{M}}_{abc}$. Our strategy moving forward will be to hold fixed the values for which the solution is nonlinear, while optimizing the parameters for which it is linear. In the case of $\boldsymbol{R}$ and $\overset{\boldsymbol{SE}}{\mathcal{M}}_{abc}$, recursion is not necessarily required, although holding $\overset{\boldsymbol{SE}}{\mathcal{M}}_{abc}$ fixed

might help improve the estimates of $\boldsymbol{R}$ slightly. We will discover in section 3.3.4.2 that recursion is actually required for the estimation of an additional parameter, $\boldsymbol{V}$, which scales the state variables so they can interact meaningfully.

3.3 Solving for the Nonlinear Parameters

The task of estimating the parameters of the system described in equation (3.22) can also be described as an optimization problem. Formally, we choose to represent it as the minimization:

$$\underset{\substack{\boldsymbol{SE}\\ \mathcal{M}}}{\arg\min} \quad E\left(\left|\left|\left(\overset{\boldsymbol{SE}}{\hat{\boldsymbol{x}}}_k - \left(\hat{\boldsymbol{R}}\overset{\boldsymbol{SE}}{\hat{\boldsymbol{x}}}_{k-1} + \sum_{abc}\left(\hat{\boldsymbol{R}}_a \overset{\boldsymbol{SE}}{\mathcal{M}}_{abc}(\overset{\boldsymbol{SE}}{\hat{\boldsymbol{x}}}_{k-1})_c(\overset{\boldsymbol{SE}}{\hat{\boldsymbol{x}}}_{k-1})_b\right)\right)\right)\right|\right|^2\right) \tag{3.23}$$

Note we continue minimizing the MSE in the hopes of leveraging the expedient and often robust framework of OLS as described in section 3.1.2. Recall that in order to apply this technique, we will need to hold certain approximate objects fixed as we optimize others.

In the previous section, we determined a method for getting an initial estimate of the linear terms and parameters. Subsequently, we determined a method for converting our nonlinear system into a linear-in-the-parameters approximation. We can now substitute the unknown linear terms and parameters for those initial estimates, and given those, derive a method for estimating the nonlinear parameters. In order to estimate the nonlinear parameters, we must first determine the nonlinear terms, which will be nonlinear functions of the linear terms $\boldsymbol{x}$. To solve for any of these unknowns will require us to examine the linear residual in closer detail.

3.3.1 Decomposition of the Linear Residual

The linear residual, $\boldsymbol{e}_{\text{lin}}$, contains components which result from various failures of the linear approximation of the system to explain the dynamics of the system at large. Rearranging the difference equation (3.22) results in a convenient form of the residual in terms of the variables we will hold constant.

$$\overset{\boldsymbol{SE}}{\hat{\boldsymbol{x}}}_k - \hat{\boldsymbol{R}} \overset{\boldsymbol{SE}}{\hat{\boldsymbol{x}}}_{k-1} \approx \sum_{abc} \left(\hat{\boldsymbol{R}}_a \overset{\boldsymbol{SE}}{\mathcal{M}}_{abc} (\overset{\boldsymbol{SE}}{\hat{\boldsymbol{x}}}_{k-1})_c (\overset{\boldsymbol{SE}}{\hat{\boldsymbol{x}}}_{k-1})_b \right) \tag{3.24}$$

$$\overset{\boldsymbol{SE}}{\boldsymbol{e}}_{\text{lin}\,k} = \overset{\boldsymbol{SE}}{\hat{\boldsymbol{x}}}_k - \hat{\boldsymbol{R}} \overset{\boldsymbol{SE}}{\hat{\boldsymbol{x}}}_{k-1} \tag{3.25}$$

Above, we have omitted the noise term $\boldsymbol{\xi}$ for clarity. As the orthography would suggest, this residual is related by a change of basis to the linear residual in the measurements space, $\boldsymbol{e}_{\text{lin}}$, which is calculated via:

$$\boldsymbol{e}_{\text{lin}} = \boldsymbol{Y}_k - \hat{\boldsymbol{A}} \boldsymbol{Y}_{k-1} \tag{3.26}$$

Notice how in this form, the scale is preserved. This turns out to be an essential property when calculating an estimate of model fitness and the contribution to fitness that each successive refinement yields. If we cannot reduce the variance of $\boldsymbol{e}_{\text{lin}}$ with subsequent nonlinear approximations of the system, then perhaps the system is dominated by linear dynamics and noise.

Meanwhile, we continue to isolate the unknown terms to result in a form compatible with the OLS construction described in section 3.1.2. To do so, we must inverse

filter the residual vector timeseries with our estimate of $\boldsymbol{R}$:

$$\overset{\boldsymbol{SE}}{\boldsymbol{e}}_k \approx \sum_{abc} \hat{\boldsymbol{R}}_a \overset{\boldsymbol{SE}}{\mathcal{M}}_{abc} (\overset{\boldsymbol{SE}}{\hat{\boldsymbol{x}}}_{k-1})_c (\overset{\boldsymbol{SE}}{\hat{\boldsymbol{x}}}_{k-1})_b \tag{3.27}$$

$$\hat{\boldsymbol{R}}^{-1} \overset{\boldsymbol{SE}}{\boldsymbol{e}}_k \approx \sum_{bc} \overset{\boldsymbol{SE}}{\mathcal{M}}_{abc} (\overset{\boldsymbol{SE}}{\hat{\boldsymbol{x}}}_{k-1})_c (\overset{\boldsymbol{SE}}{\hat{\boldsymbol{x}}}_{k-1})_b \tag{3.28}$$

This inverse filtering step provides us with another opportunity for failure, as any errors in the estimation of $\boldsymbol{R}$ will certainly bubble up into our decomposition of the linear residual. For this reason, our estimate of the linear component must be robust. While LPC may be suitable for directly estimating the linear component of the state update equations, other techniques may be more appropriate when the system under study is sufficiently nonlinear. Furthermore, a parametric approximation of the order of the linear basis may be inappropriate for certain applications. We discuss the ramifications of these cases in section 3.4.

3.3.2 Nonlinear Least Squares

The form of equation (3.28) is sufficiently close to the linear-in-the-parameters form first described in section 3.1.2 that we can derive a non-recursive construction of this stage of the optimization simply by permuting some of the coefficients. Once again, we will need to convert from an underdetermined system with infinitely many solutions, to an overdetermined system with no solution. This time, however, rather than solving a single system of N simultaneous equations, we must solve N systems of N simultaneous equations.

3.3.2.1 The Underdetermined Form

We now form the equation for the state vector in terms of a nonlinear function of the previous state vector $\tilde{\boldsymbol{p}}$, an unknown matrix of parameters $\tilde{\Theta}$, and the nonlin-

ear residual $\tilde{e}$. We intentionally use a notation as visually reminiscent as possible to the underdetermined form of the linear-in-the-parameters model described in equation (3.1):

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{p}}\tilde{\Theta} + \tilde{e} \tag{3.29}$$

given the following definitions:

$$\tilde{\boldsymbol{y}} \equiv \left(\hat{\boldsymbol{R}}^{-1} \boldsymbol{SE}_{\text{lin}\,k} \right)^{\text{T}} \tag{3.30}$$

$$\tilde{\boldsymbol{p}} \equiv (q_1 \dots q_{N^2}) \tag{3.31}$$

$$\tilde{\Theta} \equiv \begin{pmatrix} \mu_{111} & \cdots & \mu_{N11} \\ \vdots & \ddots & \\ \mu_{1NN} & & \mu_{NNN} \end{pmatrix} \tag{3.32}$$

That the term coefficients q populate the term vector $\boldsymbol{p}$ is a consequence of our choice in notation for the nonlinear transformation on the terms, which will be derived shortly in section 3.3.2.2.

Notice that the dimensions of the multilinear regression problem we have set up require that the left-hand side, here notated as $\tilde{\boldsymbol{y}}$, be transposed. The column space of this system represents each independent system of N equations which $\tilde{\Theta}$ must solve.

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{p}}\tilde{\Theta} + \tilde{\boldsymbol{e}} \begin{cases} \tilde{y}_1 = \tilde{\boldsymbol{p}}\tilde{\theta}_{(:1)} + \tilde{\boldsymbol{e}} \\ \tilde{y}_2 = \tilde{\boldsymbol{p}}\tilde{\theta}_{(:2)} + \tilde{\boldsymbol{e}} \\ \vdots \\ \tilde{y}_N = \tilde{\boldsymbol{p}}\tilde{\theta}_{(:N)} + \tilde{\boldsymbol{e}} \end{cases} \tag{3.33}$$

3.3.2.2 Nonlinear Transformation of Model Terms

As mentioned in the previous section, we must define a nonlinear transformation on the state vector $\boldsymbol{x}$ that isolates the nonlinear action of our linear-in-the-parameters approximation of the system. Referring to the right-hand side of the residual definition in equation (3.28), we find the terms enter as a homogeneous quadratic in N variables:

$$\boldsymbol{Q}_{k(a,b)} \equiv (\overset{\boldsymbol{SE}}{\hat{x}}_{ak})(\overset{\boldsymbol{SE}}{\hat{x}}_{bk}) \tag{3.34}$$

To simplify the notation further, we apply a vector-valued indexing function along the second index of $\boldsymbol{Q}$. In our definition, we state that the row space of $\boldsymbol{Q}$ corresponds to each time step.

$$\boldsymbol{Q} \equiv \begin{pmatrix} \overset{\boldsymbol{SE}}{x}_{11}\overset{\boldsymbol{SE}}{x}_{11} & \overset{\boldsymbol{SE}}{x}_{11}\overset{\boldsymbol{SE}}{x}_{21} & \cdots & \cdots & \overset{\boldsymbol{SE}}{x}_{N1}\overset{\boldsymbol{SE}}{x}_{N1} \\ \overset{\boldsymbol{SE}}{x}_{12}\overset{\boldsymbol{SE}}{x}_{12} & \overset{\boldsymbol{SE}}{x}_{12}\overset{\boldsymbol{SE}}{x}_{22} & \cdots & \cdots & \overset{\boldsymbol{SE}}{x}_{N2}\overset{\boldsymbol{SE}}{x}_{N2} \\ \vdots & \vdots & & & \vdots \\ \overset{\boldsymbol{SE}}{x}_{1K}\overset{\boldsymbol{SE}}{x}_{1K} & \overset{\boldsymbol{SE}}{x}_{1K}\overset{\boldsymbol{SE}}{x}_{1K} & \cdots & \cdots & \overset{\boldsymbol{SE}}{x}_{NK}\overset{\boldsymbol{SE}}{x}_{NK} \end{pmatrix} \tag{3.35}$$

Practically speaking, the entire matrix $\boldsymbol{Q}$ need not be computed. If each row is mapped into a quadratic matrix which results from the outer product of $\overset{\boldsymbol{SE}}{\boldsymbol{x}}$, only the upper or lower triangular region need be calculated. If the matrix is to be constructed and held entirely in memory (not recommended for large scale computations), the remaining coefficients may be copied from their redundant locations.

Also note that since these terms are quadratics, $\boldsymbol{Q}$ will not be mean zero. It turns out that this is not required for the machinery of OLS to function properly in this context. In fact, removing the offset to center the signals about 0 results in erroneous calculations of $\overset{\boldsymbol{SE}}{\hat{\mathcal{M}}}$, and should therefore be avoided.

Finally, we note that a second estimate of $\boldsymbol{R}$ could be derived by simply adding the appropriate number of columns to $\boldsymbol{Q}$ with $\hat{\boldsymbol{x}}$ in each, however this is not our concern presently. We will discuss such an extension to the technique in section 5.2.

3.3.2.3 The Overdetermined Form

Our next task is to form a block of the auto-covariance estimate matrix of quadratic state variable terms, which itself is comprised of quartic terms:

$$\mathcal{G}_{(ab)(cd)} \equiv \sum_K \boldsymbol{Q}_{k(a,b)} \boldsymbol{Q}_{k(c,d)} \tag{3.36}$$

$$= \sum_K \begin{pmatrix} \boldsymbol{Q}_{k(1,1)}\boldsymbol{Q}_{k(1,1)} & \boldsymbol{Q}_{k(1,1)}\boldsymbol{Q}_{k(1,2)} & \cdots & \cdots & \boldsymbol{Q}_{k(1,1)}\boldsymbol{Q}_{k(N,N)} \\ \boldsymbol{Q}_{k(1,2)}\boldsymbol{Q}_{k(1,1)} & \boldsymbol{Q}_{k(1,2)}\boldsymbol{Q}_{k(1,2)} & \cdots & \cdots & \boldsymbol{Q}_{k(1,2)}\boldsymbol{Q}_{k(N,N)} \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & \\ \boldsymbol{Q}_{k(N,N)}\boldsymbol{Q}_{k(1,1)} & \boldsymbol{Q}_{k(N,N)}\boldsymbol{Q}_{k(1,2)} & \cdots & \cdots & \boldsymbol{Q}_{k(N,N)}\boldsymbol{Q}_{k(N,N)} \end{pmatrix} \tag{3.37}$$

The same considerations regarding the redundancy of this construction as were observed in the construction of equation (3.35) apply here. We once again leverage a vector-valued index, this time along both indices of $\mathcal{G}$.

These N^2xN^2 blocks are then copied into the N^3xN^3 block-wise auto-covariance estimate matrix $\boldsymbol{G}$, along the diagonal. This redundant structure reflects the fact that we

are solving N systems of N dimensions each:

$$\boldsymbol{G} = \begin{bmatrix} \mathcal{G} & & & \\ & \mathcal{G} & & \\ & & \ddots & \\ & & & \mathcal{G} \end{bmatrix} \tag{3.38}$$

$$\boldsymbol{G}_{(abc)(def)} = \begin{bmatrix} g_{(111)(111)} & g_{(111)(112)} & \cdots & \cdots & \cdots & g_{(111)(NNN)} \\ g_{(112)(111)} & g_{(112)(112)} & \cdots & \cdots & \cdots & g_{(112)(NNN)} \\ \vdots & \vdots & & & & \vdots \\ \vdots & \vdots & & & & \vdots \\ \vdots & \vdots & & & & \vdots \\ g_{(NNN)(111)} & g_{(NNN)(112)} & \cdots & \cdots & \cdots & g_{(NNN)(NNN)} \end{bmatrix} \tag{3.39}$$

Both indices into $\boldsymbol{G}$ are 3-vectors of dimension $NxNxN$. It is worth noting here that if a single block $\mathcal{G}$ were invertible, we could greatly reduce the complexity of the algorithm. We will turn back to this idea in section 3.3.3.

We must now define a function to map the cross-covariance estimates of the nonlinearly transformed previous state variables with the current inverse filtered residual. The most challenging aspect of this step in practice is the task of keeping indices properly

ordered.

$$f_{(abc)} = \sum_{Kd} \left(r^{-1}\right)_{ad} \overset{\boldsymbol{SE}}{\hat{e}}_{dk} \boldsymbol{Q}_{k(bc)} \tag{3.40}$$

$$\boldsymbol{f} = \begin{pmatrix} f_{(111)} \\ f_{(112)} \\ \vdots \\ \vdots \\ f_{(NNN)} \end{pmatrix} \tag{3.41}$$

Just as in our notation for $\boldsymbol{G}$ in the derivation above in equation (3.38), we use a 3-vector notation for indexing into the row dimension of this vector.

Finally, we are poised to suggest a multilinear regression equation for the solution of the coefficients in $\overset{\boldsymbol{SE}}{\hat{\mathcal{M}}}$:

$$\overset{\boldsymbol{SE}}{\hat{\mu}}_{abc} = \left(\boldsymbol{G}^{-1}\boldsymbol{F}\right)_{(abc)} \tag{3.42}$$

$$= \sum_{def} g^{-1}_{(abc)(def)} f_{(def)} \tag{3.43}$$

Unfortunately, however, this suggestion is inadequate as $\boldsymbol{G}$ is often singular, and in general very poorly conditioned. Once again, we find ourselves with no solutions, despite having formed the estimates in precisely the same fashion as the linear estimates in section 3.1.2.

3.3.3 Linear-in-the-Parameters Optimization with Constraints

In the previous section, we discussed the possible application of a multilinear regression solution to the normal equations of our linear-in-the-parameters system approximation. This solution is unsatisfactory because equation (3.42) is poorly conditioned in

general. In this section we discuss a principled technique for the efficient solution of this step of the algorithm.

3.3.3.1 Degeneracy of the Overdetermined Form

In our derivation, $\boldsymbol{G}$ is in general nearly singular, as are the diagonal blocks $\mathcal{G}$ that comprise it. This is because the solutions to this system, $\hat{\mathcal{M}}^{\boldsymbol{SE}}$, reside in a subspace of $\mathcal{M}$.

3.3.3.2 Finding a Basis for a Subspace of $\mathcal{M}$

To form a basis for the subspace of controllable and observable choices of $\mathcal{M}^{\boldsymbol{SE}}$, we choose sparse 3-tensors which satisfy the constraints (2.31) and (2.32). We access the components of each tensor using a 3-vector in a manner identical to equation (3.40), thus treating each tensor as if it is a vector of dimension N^3. These emerge in two subtypes: those with either two or three unique nonzero indices. Vectors specified by 2-permutations of indices may be constructed by the following function:

$$\mathbf{\Pi}(\mathrm{ab};1)_{def} = \begin{cases} 1 & \mathrm{a} \neq \mathrm{b} \\ 1 & \mathrm{d} = \mathrm{a}, \mathrm{e} = \mathrm{f} = \mathrm{b} \\ -1 & \mathrm{e} = \mathrm{a}, \mathrm{d} = \mathrm{f} = \mathrm{b} \\ 0 & \text{otherwise} \end{cases} \tag{3.44}$$

Where choices of a and b specify each permutation of 2 indices permitted by the dimensionality of $\mathcal{M}$. We might choose the 3-tensors satisfying (2.31) and (2.32) with three

unique indices def to take the form:

$$\mathrm{a} \neq \mathrm{b} \neq \mathrm{c}$$

$$\mathbf{\Pi}(\mathrm{abc};1)_{def} = \begin{cases} 1 & (\mathrm{def}) = (\mathrm{abc}), \text{ or } (\mathrm{def}) = (\mathrm{cab}) \\ -1 & (\mathrm{def}) = (\mathrm{bac}), \text{ or } (\mathrm{def}) = (\mathrm{acb}) \\ 0 & \text{otherwise} \end{cases} \tag{3.45}$$

Although the 3-tensors resulting from functions (3.44) and (3.45) span the subspace of observable and controllable $\overset{\boldsymbol{SE}}{\mathcal{M}}$, they do not form a basis of the subspace. An orthonormal basis of this subspace may be defined as follows. Given $\mathrm{a} \neq \mathrm{b}$:

$$\mathbf{\Pi}(\mathrm{ab};1)_{def} = \begin{cases} \frac{1}{\sqrt{2}} & \mathrm{d} = \mathrm{a}, \mathrm{e} = \mathrm{f} = \mathrm{b} \\ -\frac{1}{\sqrt{2}} & \mathrm{e} = \mathrm{a}, \mathrm{d} = \mathrm{f} = \mathrm{b} \\ 0 & \text{otherwise} \end{cases} \tag{3.46}$$

We have applied a scaling factor defined by $\left(\sqrt{\sum_{N^3}\left(\mathbf{\Pi}(\mathrm{ab};1)_{def}\right)^2}\right)^{-1}$.

Given $\mathrm{a} < \mathrm{b} < \mathrm{c}$, the remaining orthonormal basis vectors are:

$$\mathbf{\Pi}(\mathrm{abc};1)_{def} = \begin{cases} \frac{1}{2} & (\mathrm{def}) = (\mathrm{abc}), \text{ or } (\mathrm{def}) = (\mathrm{cab}) \\ -\frac{1}{2} & (\mathrm{def}) = (\mathrm{bac}), \text{ or } (\mathrm{def}) = (\mathrm{acb}) \\ 0 & \text{otherwise} \end{cases} \tag{3.47}$$

$$\mathbf{\Pi}(\mathrm{abc};2)_{def} = \frac{1}{\sqrt{12}}\Big(\mathbf{\Pi}(\mathrm{bca};1) - \mathbf{\Pi}(\mathrm{cab};1)\Big) \tag{3.48}$$

We recognize that $\mathbf{\Pi}$ is by no means unique. In fact, we could add coefficients from any totally antisymmetric tensor to our vectors and observe no functional change in

the resulting analysis. However, this basis is one whose nullspace spans the subspace of totally antisymmetric tensors. Even in this quality $\mathbf{\Pi}$ is not unique. There are 6 choices of bases sufficient for describing possible coordinate systems which have an antisymmetric component of zero. We could indeed have chosen a different basis– e.g. $\mathbf{\Pi}(\text{bca};1)$– in equation (3.47). An additional set of alternative orthogonal bases could be derived by negating the components of $\mathbf{\Pi}$ or its alternatives. Numerically, we could have found this basis by applying Gram-Schmidt to a basis spanning the nullspace of totally antisymmetric tensors. However, since this basis is only a function of the dimensionality of $\mathcal{M}$, and not dependent on the data itself, we only need to solve it once for a given system dimension. Furthermore, the orthogonalization step in certain implementations of Gram-Schmidt is notoriously unstable. We therefore opt for a general, closed-form solution.

Using this basis, we project $\overset{\boldsymbol{SE}}{\mathcal{M}}$ into the subspace as follows:

$$\overset{\boldsymbol{SE}}{\mathcal{M}'}_d = \sum_{abc}^{N} \left(\mathbf{\Pi}_{d,(abc)} \overset{\boldsymbol{SE}}{\mathcal{M}}_{abc} \right) \tag{3.49}$$

We similarly project $\boldsymbol{G}$ and $\boldsymbol{F}$ into the subspace:

$$\boldsymbol{G'} = \mathbf{\Pi} \boldsymbol{G} \mathbf{\Pi}^{\mathrm{T}} \tag{3.50}$$

$$\boldsymbol{F'} = \mathbf{\Pi} \boldsymbol{F} \tag{3.51}$$

The dimensionality of this subspace is $2\left(\binom{N}{2} + \binom{N}{3}\right)$ or $N(N^2-1)/3$.

3.3.3.3 Solving for $\mathcal{M}$ within a Well-Conditioned Subspace

Now that we have transformed our poorly conditioned N^3 dimensional system into a well conditioned system of dimension $N(N^2-1)/3$, we can proceed to solve for

the parameter tensor $\mathcal{M}'$ like so:

$$\overset{\boldsymbol{SE}}{\hat{\mathcal{M}}'}_d = \boldsymbol{G'}^{-1}\boldsymbol{F'} \tag{3.52}$$

We can then recover the resulting estimate of $\mathcal{M}$ with the operation:

$$\overset{\boldsymbol{SE}}{\hat{\mathcal{M}}}_{abc} = \left(\boldsymbol{\Pi}^{\mathrm{T}}\overset{\boldsymbol{SE}}{\hat{\mathcal{M}}'}\right)_{(abc)} \tag{3.53}$$

The operation in equation (3.53) applies our structural constraints that $\overset{\boldsymbol{SE}}{\mathcal{M}}$ must satisfy onto any estimate $\overset{\boldsymbol{SE}}{\hat{\mathcal{M}}}$. This operation will minimize the total distance, orthogonal to the space of acceptable $\hat{\mathcal{M}}$, between the components of its input and its resulting constrained $\overset{\boldsymbol{SE}}{\hat{\mathcal{M}}}$. Note that minimizing the MSE in modal space does not guarantee an optimal parameter estimate in the space of the measurements.

3.3.4 Determining a Norm for $\mathcal{M}$

The regression defined in equation (3.52) will result in the unique estimate of $\mathcal{M}'$ that minimizes the MSE of the state space. In the case where $\boldsymbol{W}$ is orthogonal, this error measure is sufficient to describe the fitness to the measured output characteristics of the system. In the more general case, such an assumption is restrictive. This is because pairs of columns of $\boldsymbol{S}$ may be scaled and rotated by arbitrary complex numbers while $\boldsymbol{S}$ retains the property that it diagonalizes the linear state transition matrix $\boldsymbol{T}$. Therefore, the power in a given modal state variable pair can be estimated down to a scale factor. The relative power of each mode, and the phase relationships between those modes, are indeterminate, arbitrary and unobservable in the linear case. However, in the case where nonlinear coupling may be present, the power relationships between modal pairs are vital

to determining $\overset{\boldsymbol{SE}}{\mathcal{M}}$, whose off-diagonal blocks of 2*x*2*x*2 encode interactions between modes.

We could define $\boldsymbol{V}$ in terms of the relationship between the eigenvector basis $\boldsymbol{S}$ and $\boldsymbol{W}$, assuming knowledge of the latter:

$$\hat{\boldsymbol{S}} = \boldsymbol{V}^{-1}\overset{\boldsymbol{S}}{\boldsymbol{W}} \tag{3.54}$$

where $\boldsymbol{V}$ is a diagonal, square matrix with complex conjugates down the diagonal. This relation assumes that $\boldsymbol{W}$ is square, so that operations are defined. In general, this relation will hold for each row in $\boldsymbol{S}$ and $\overset{\boldsymbol{S}}{\boldsymbol{W}}$.

3.3.4.1 The Energy Basis

The diagonal matrix $\boldsymbol{V}$ rescales the variables in the eigenbasis such that each mode is represented in the same units. If we apply this rescaling in the real-valued block diagonalized basis, the resulting state variables represent the instantaneous energy states of each of the system's degrees of freedom:

$$\boldsymbol{x} = \boldsymbol{SE}\overset{\boldsymbol{E}}{\boldsymbol{V}}\overset{\boldsymbol{SE}}{\boldsymbol{x}}\overset{\boldsymbol{E}}{\boldsymbol{V}} \tag{3.55}$$

Applying this change of basis in the usual fashion and propagating it throughout the approximated state update equations (3.22) results in:

$$\overset{\boldsymbol{SE}}{\boldsymbol{x}}\overset{\boldsymbol{E}}{\boldsymbol{V}}_k \approx \left(\overset{\boldsymbol{E}}{\boldsymbol{V}}\right)^{-1}\boldsymbol{R}\overset{\boldsymbol{E}}{\boldsymbol{V}}\overset{\boldsymbol{SE}}{\boldsymbol{x}}\overset{\boldsymbol{E}}{\boldsymbol{V}}_{k-1} + \sum_{abc}\left(\left(\left(\overset{\boldsymbol{E}}{\boldsymbol{V}}\right)^{-1}\boldsymbol{R}\overset{\boldsymbol{E}}{\boldsymbol{V}}\left(\overset{\boldsymbol{E}}{\boldsymbol{V}}\right)^{-1}\right)_a \overset{\boldsymbol{SE}}{\mathcal{M}}_{abc}\left(\overset{\boldsymbol{E}}{\boldsymbol{V}}\overset{\boldsymbol{SE}}{\boldsymbol{x}}\overset{\boldsymbol{E}}{\boldsymbol{V}}_{k-1}\right)_b\left(\overset{\boldsymbol{E}}{\boldsymbol{V}}\overset{\boldsymbol{SE}}{\boldsymbol{x}}\overset{\boldsymbol{E}}{\boldsymbol{V}}_{k-1}\right)_c\right) \tag{3.56}$$

We have dropped the noise term for clarity above. It is: $\overset{\boldsymbol{SE}\overset{\boldsymbol{E}}{\boldsymbol{V}}}{\boldsymbol{\xi}_k} = \left(\overset{\boldsymbol{E}}{\boldsymbol{V}}\right)^{-1} \boldsymbol{E}^* \boldsymbol{S}^{-1} \boldsymbol{\xi}_k$. Finally, we have a definition of how $\mathcal{M}$ transforms with $\boldsymbol{V}$:

$$\overset{\boldsymbol{SE}\overset{\boldsymbol{E}}{\boldsymbol{V}}}{\mathcal{M}_f} \equiv \left(\left(\overset{\boldsymbol{E}}{\boldsymbol{V}}\right)^{-1} \boldsymbol{E}^* \boldsymbol{S}^{-1}\right) \mathcal{M}_c \left(\boldsymbol{SE}\overset{\boldsymbol{E}}{\boldsymbol{V}}\right) \left(\boldsymbol{SE}\overset{\boldsymbol{E}}{\boldsymbol{V}}\right)_{fc} \tag{3.57}$$

The metric which defines the complex-valued energy basis can be decomposed into its polar coordinates for further analysis. The relative magnitudes of $\boldsymbol{V}$ have an impact on the measured outputs of the system, whereas the relative phases simply perform a rotation of the basis vectors, leaving the resulting system dynamics unaffected in the measurements. Meanwhile, in the real-valued energy basis, $\overset{\boldsymbol{E}}{\boldsymbol{V}}$ is block-diagonal:

$$\overset{\boldsymbol{E}}{\boldsymbol{V}} = \begin{bmatrix} \begin{pmatrix} a_1 & b_1 \\ -b_1 & a_1 \end{pmatrix} & & \\ & \ddots & \\ & & \begin{pmatrix} a_{N/2} & b_{N/2} \\ -b_{N/2} & a_{N/2} \end{pmatrix} \end{bmatrix} \tag{3.58}$$

Since $\boldsymbol{W}$ is generally unknown, the relation described in equation (3.54) is merely a theoretical construction. In practice, $\boldsymbol{V}$ must be estimated given only measurement terms and the parameters we have estimated thus far.

3.3.4.2 Solving for $\boldsymbol{V}$

At this point, we have found a biased estimate of $\hat{\mathcal{M}}$ using a low-order approximation of the model. However, there are differences between this simplified form and the form we presented in chapter 2, with an upper bound described in equation 3.16. This low-order approximation has served its purpose. We now return back to the exponential

form as we continue refining the approximation. The exponential form of the biased resynthesis error is:

$$\begin{aligned} \boldsymbol{e}_{\exp k} &= \hat{\boldsymbol{S}}\boldsymbol{E}\left(\overset{\hat{\boldsymbol{S}}\boldsymbol{E}}{\boldsymbol{x}}_k - \hat{\boldsymbol{R}}\exp\left(\overset{\boldsymbol{S}\boldsymbol{E}}{\hat{\mathcal{M}}}\overset{\hat{\boldsymbol{S}}\boldsymbol{E}}{\boldsymbol{x}}_{k-1}\right)\overset{\hat{\boldsymbol{S}}\boldsymbol{E}}{\boldsymbol{x}}_{k-1}\right) \\ &= \hat{\boldsymbol{S}}\boldsymbol{E}\left(\overset{\hat{\boldsymbol{S}}\boldsymbol{E}}{\boldsymbol{x}}_{k|k} - \overset{\hat{\boldsymbol{S}}\boldsymbol{E}}{\boldsymbol{x}}_{k|k-1}\right) \end{aligned} \tag{3.59}$$

It may be helpful to consider the biased resynthesis error to be a special case where $\hat{\boldsymbol{V}}$ is the identity matrix. The goal of this stage of the parameter estimation is to minimize the resynthesis error by removing this bias in $\hat{\boldsymbol{V}}$, which is an artifact of our sensor placement, and not an intrinsic property of the system.

Effectively, we need only find the relative weights to apply to each of the modes that minimizes the resynthesis error. The optimization problem can be framed as:

$$\underset{\overset{\boldsymbol{E}}{\hat{\boldsymbol{V}}}}{\arg\min}\left(E\left(\left\|\hat{\boldsymbol{S}}\boldsymbol{E}\overset{\boldsymbol{E}}{\hat{\boldsymbol{V}}}\left(\overset{\hat{\boldsymbol{S}}\boldsymbol{E}\overset{\boldsymbol{E}}{\hat{\boldsymbol{V}}}}{\boldsymbol{x}}_k - \hat{\boldsymbol{R}}\exp\left(\overset{\boldsymbol{S}\boldsymbol{E}\overset{\boldsymbol{E}}{\hat{\boldsymbol{V}}}}{\hat{\mathcal{M}}}\overset{\hat{\boldsymbol{S}}\boldsymbol{E}\overset{\boldsymbol{E}}{\hat{\boldsymbol{V}}}}{\boldsymbol{x}}_{k-1}\right)\overset{\hat{\boldsymbol{S}}\boldsymbol{E}\overset{\boldsymbol{E}}{\hat{\boldsymbol{V}}}}{\boldsymbol{x}}_{k-1}\right)\right\|^2\right)\right) \tag{3.60}$$

We propose an iterative solution, which tests successive candidates of $\boldsymbol{V}$ in the pattern of a breadth-first binary search.

Although equation (3.57) describes how to transform $\overset{\boldsymbol{S}\boldsymbol{E}}{\mathcal{M}}$ with a given choice of $\hat{\boldsymbol{V}}$, this definition assumes that such a transformation is orthogonal. Unfortunately, orthogonal transformations in the energy basis have no effect on the resynthesis error. The residual error is also invariant to global scale factors applied to $\hat{\boldsymbol{V}}$. Furthermore, the coefficients appear in pairs corresponding to the modal variables. Because of these structural invariants, we can safely constrain our search to the positive reals, in general only searching for $(N/2)-1$ free parameters.

3.3.5 Cross-Model Error Measures

It is clear from the above discussion that the parameters we wish to estimate are not unique. This poses some challenges if we wish to compare our estimates to either ground truth, earlier estimates, or other systems entirely. For this reason, we require the definition of principled error measures which can be compared without recourse to the larger context of the model that yields them.

The RMS of the nonlinear resynthesis error $E\left(||\boldsymbol{e}_{\text{exp}}||^2\right)$ is a poor approximation of overall parametric fit for a number of reasons. First, it cannot be used for comparison across different models if those models have different variances. Moreover, $E\left(||\boldsymbol{e}_{\text{exp}}||^2\right)$ contains contributions from $\boldsymbol{\xi}$, which is intended to describe any dynamics the model fails to explain. In general, a perfectly accurate parameter estimate should not cause a measure like $E\left(||\boldsymbol{e}_{\text{exp}}||^2\right)$ to disappear completely. The lower bound on this measure is $E\left(||\boldsymbol{W}\boldsymbol{\xi}^{\text{T}}||^2\right)$. In a simulation context, where directly calculated accuracy assessments are required, this component could be removed from the resynthesis error term to result in the purely deterministic component of the resynthesis error.

3.3.5.1 Measuring incremental improvements in fitness

In section 3.1.2, we discussed the framework for OLS as an optimization technique via an error function:

$$E\left(||\boldsymbol{y}-\boldsymbol{P}\hat{\theta}||^2\right)=E\left(||\boldsymbol{Y}_k-\hat{\boldsymbol{A}}\boldsymbol{Y}_{k-1}||^2\right) \tag{3.61}$$

This error function is used similarly to that of the nonlinear OLS stage of the algorithm, described in section (3.23), except that the nonlinear MSE is not specified in the measurement space. Thus, the true nonlinear counterpart to equation (3.61) is equation (3.60), which incorporates the complete transform from state space to measurement space.

The ratio between linear and nonlinear MSE gives us a measure of the improvement in variance between the linear and nonlinear estimates:

$$e_\Delta \equiv 20\log_{10}\left(\frac{E\left(||\boldsymbol{e}_{\text{lin}}||^2\right)}{E\left(||\boldsymbol{e}_{\text{exp}}||^2\right)}\right) \tag{3.62}$$

We will use this measure to determine the benefit afforded us by the additional complexity of the nonlinear parameter estimation. Since the estimation of the nonlinear parameters is recursive in nature, the measure $e_{\Delta a}$ can be calculated repeatedly for each $\boldsymbol{e}_{\text{exp}a}$ which accompanies each recursion.

3.3.5.2 Index of Nonlinearity

While the improvement metric e_Δ offers insight about the performance of the algorithm across various stages of analysis, it does not provide a measurement of an intrinsic property of the model. Another dimensionless quantity may be derived using similar means which provides such a cross-model description of the overall nonlinearity of the model in question:

$$\Phi \equiv 20\log_{10}\left(\frac{E\left(\left|\left|\boldsymbol{SEV}\left(\boldsymbol{R}\exp\left(\begin{matrix}\boldsymbol{SEV}\\ \mathcal{M}\end{matrix}\boldsymbol{SEV}_{\boldsymbol{x}\,1:k-1}\right)\boldsymbol{SEV}_{\boldsymbol{x}\,1:k-1}-\boldsymbol{R}\,\boldsymbol{SEV}_{\boldsymbol{x}\,1:k-1}\right)-E(\boldsymbol{y})\right|\right|^2\right)}{E\left(||\boldsymbol{y}_{2:k}-E(\boldsymbol{y})||^2\right)}\right) \tag{3.63}$$

This dimensionless coefficient is an unbiased measure of the portion of the total variance attributed to nonlinear dynamics. Note that equation (3.63) implicitly distinguishes among linear, nonlinear, and stochastic components of the output. In practice, when the only term given is $\boldsymbol{Y}$, we can still calculate an approximate $\hat{\Phi}$.

3.4 Considerations for Rectangular $\boldsymbol{W}$

The above derivation assumed $\boldsymbol{W}$ was square and invertible. When the system exhibits a different number of modes than the number of sensors used to capture its dynamics, $\boldsymbol{W}$ is rectangular. It may often happen that the modes outnumber the sensor channels. In this case, it is often difficult to estimate the linear parameters from the measurements if the system is sufficiently nonlinear.

One aspect of the derivation that breaks down in this case is equation (3.54), which we can solve for $\boldsymbol{V}$:

$$v_{ii} = \frac{1}{N}\sum_{j}^{N}\left(\frac{\left(\boldsymbol{W}\boldsymbol{E}^{\mathrm{T}}\right)_{ji}}{\boldsymbol{S}_{ji}}\right), \qquad 0 \quad \text{otherwise} \tag{3.64}$$

when $\boldsymbol{W}$ is square. In fact, in the ideal case, we notice that each coefficient in the component-wise quotient $\left(\boldsymbol{W}\boldsymbol{E}^{\mathrm{T}}\right)_{ji}/\boldsymbol{S}_{ji}$ is constant along j. Each row of $\boldsymbol{W}$ effectively describes the mixture of modal signals present in each sensor channel, presumably as a result of spatial orientation of the sensors. We assume these mixtures remain constant. Since the columns of $\boldsymbol{S}$ each describe a linear evolution in time, we can expect to see powers of $\boldsymbol{\Lambda}$ appear in blocks of $\boldsymbol{S}$ corresponding to simultaneous measurements.

We can leverage this Vandermonde-like structure to estimate $\boldsymbol{V}$ when $\boldsymbol{W}$ is rectangular:

$$v_{ii} = \frac{1}{N}\sum_{j}^{N}\frac{\begin{bmatrix}\left(\boldsymbol{W}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{\Lambda}^{0}\right)\\ \left(\boldsymbol{W}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{\Lambda}^{1}\right)\\ \vdots \\ \left(\boldsymbol{W}\boldsymbol{E}^{\mathrm{T}}\boldsymbol{\Lambda}^{L}\right)\end{bmatrix}_{ji}}{\boldsymbol{S}_{ji}}, \qquad 0 \quad \text{otherwise} \tag{3.65}$$

where L is the number of time lags. Each block has the same dimensions as $\boldsymbol{W}$. The columns along j are once again constant, as the structure of $\boldsymbol{S}$ is compensated by the

powers of $\mathbf{\Lambda}$ in the blocks of the denominator.

The closed-form expression in equation (3.65) is of limited utility in cases where there is no principled way to estimate or measure $\boldsymbol{W}$. In field conditions, we do not know $\boldsymbol{W}$ *a priori*. However, we still can apply knowledge of $\boldsymbol{S}$'s structure to analyses where the eigenbasis is a poor estimate of the basis for the linear parameters in the system.

When analyzing sufficiently noisy or nonlinear signals, errors may build up in the regions of the eigendecomposition which correspond to lagged measurements. This happens because the lagged measurements do not contain corrections from the most current noise $\boldsymbol{\xi}$. Since the eigenvalues essentially track spectral peaks in the signal, we might imagine applying a spectral peak-finding algorithm, perhaps based on a STFT, the eigenvalues of which could be leveraged to synthesize a basis for the linear parameters:

$$\hat{\boldsymbol{S}}_{alt} \equiv \begin{bmatrix} \hat{\boldsymbol{S}}_{(1:M,:)}\hat{\mathbf{\Lambda}}^{0}_{alt} \\ \hat{\boldsymbol{S}}_{(1:M,:)}\hat{\mathbf{\Lambda}}^{1}_{alt} \\ \vdots \\ \hat{\boldsymbol{S}}_{(1:M,:)}\hat{\mathbf{\Lambda}}^{L}_{alt} \end{bmatrix} \tag{3.66}$$

where $\hat{\mathbf{\Lambda}}_{alt}$ is the diagonal eigenvalue matrix which is supplied to the algorithm.

This variant on the algorithm might also be used for updates to the nonlinear analysis when it is known that the activated modes have not changed their spatial orientation. By decoupling the later stages of analysis from an initial, potentially unstable linear prediction, this variation allows for a broader utility, with potentially greater robustness to noise.

Chapter 4

Validation and Assessment

This chapter documents several test suites conducted on hundreds of simulated systems as described in Chapter 2, and recorded systems as described in Chapter 1. The collected dataset is expansive, and we document a mere fraction of the findings here, prioritizing the questions we outlined in Chapter 1. The entire dataset will be made available in the coming months. Section 4.1 presents the assessment based on synthesized data, and section 4.2 chronicles the attempt to apply this analysis technique to field measurements, collected from the recording sessions described in section 1.1.1.

4.1 Retrieval of Parameters from Synthesized Measurements

A number of test suites were designed to validate the analysis techniques described in chapter 3. In addition to simple validation, further data was collected to calculate high-level cross-model coefficients such as Nonlinearity Index (c.f. section 3.3.5.2) and improvement metrics. Since the parameter space is quite large, the models were randomly generated within constraints for each system tested. Clearly, both param-

eters intrinsic (e.g. $\mathbf{\Lambda}$, $\mathcal{M}$, etc) and extrinsic (e.g. $\boldsymbol{W}$) to the system affect these metrics, and must be varied across models in the hopes of gathering a complete understanding of the parameter space.

For convenience, the taxonomy is as follows: Battery, (Suite,) Trial, Test. Each Battery has an associated research question. In certain cases, like Battery 1, we have multiple Suites which test aspects of a larger topic. Trials are associated with the independent variable(s) under study. Finally, we have three Tests for each system. All groupings are numbered starting with zero.

Our first battery, described in section 4.1.1, does not correspond to a research question asked at the outset of this document in section 1.3. We simply did not have a vocabulary for talking about $\mathcal{M}$, and had hoped we could prescribe Φ as a synthesis parameter. However, the implementation of such an idea turns out to be more complicated because of the multivariate nature of $\mathcal{M}$ and its interactions with $\mathbf{\Lambda}$ and $\boldsymbol{\xi P}$. We therefore prepended an additional battery, essentially exploratory in nature, to observe the interactions among these three variables.

Following this initial exploration, we continue to ask our research questions as posed in section 1.3. In short, these tested the algorithm's ability to perform as a function of sampling rate, analysis window size, and number of simultaneous measurements. Each question has a dedicated test battery, which consists of several trials. These trials are each performed on a uniquely synthesized timeseries. Unless otherwise specified, the model parameters vary randomly within ranges. By default, $\mathbf{\Lambda}$ consists of poles with an angle no greater than $\frac{\pi}{4}$ and radii no less than 0.9999, and $\mathcal{M}$ a basic tensor of magnitude 0.01. We used windows 100000 samples in length unless otherwise noted.

Generally speaking, within each trial, 3 separate tests were performed. The first test explicitly calculated the norm $\boldsymbol{V}$ from *a priori* knowledge of the measurement transformation $\boldsymbol{W}$ as in section 3.3.4, while deriving all other parameters and bases

empirically. The second calculated all parameters and bases from observations, with the eigenbasis derived via LPC as in section 3.1.3. The third assumed *a priori* knowledge of the eigenvalues, which might simulate the exogenous parameter usage described in section 3.4. Depending on the suite in which these trials were situated, certain tests may have been omitted or altered to maintain relevance to the larger context.

The results of these tests are differentiated from one another through subscripts. For example, in the case of the nonlinear exponential error measure $e_{\exp}$, the results of test 0 (with $\boldsymbol{W}$ known *a priori*) will be denoted $e_{\exp 0}$, the results of test 1 (with all bases empirically derived) will be denoted $e_{\exp 1}$, and the results of test 2 (with $\boldsymbol{S}$ known *a priori*) will be denoted $e_{\exp 2}$. An analogous notation will be used for other metrics such as Φ.

4.1.1 Battery 0: Linear Q versus Nonlinear $\left|\left|\mathcal{M}\right|\right|^2$

4.1.1.1 Methods

When the system under scrutiny is entirely linear, there is a tendency for LPC to fail if the spectral peaks are insufficiently pronounced. Since the linear component of the analysis algorithm might derive its basis from LPC, it follows that a thorough validation would investigate the effect of Q on the algorithm's accuracy.

However, if the linear filter quality is increased, the nonlinearity will also become more pronounced, because it is scale-dependent. Since we have no principled way of specifying Φ *a priori*, we first needed to find an empirical description of the relationship between $\left|\left|\mathcal{M}\right|\right|^2$, Q, and Φ.

For this battery we used 7 different systems, each with up to 16 variants : 4 different Q scale-factors, and 4 different $\mathcal{M}$ scale-factors. These were then plotted together against Φ and $e_{\exp}$ in figures 4.1 and 4.2, respectively.

4.1.1.2 Results

We can see from figure 4.1 that the most consistent increase in Φ is caused by an increase in $\left|\left|\mathcal{M}\right|\right|^2$, while the effect of Q on Φ depends entirely on the system's particulars. This makes intuitive sense: if the nonlinearity is not along a direction in the subspace also occupied by $\mathbf{\Lambda}$, then the nonlinearity index might even be inversely proportional to Q, as more of the system's behavior becomes dominated by the linear component.

Next, we varied the same two parameters and observed the change in e_{exp}. We can see from figure 4.2 that the relationship between these these two parameters and the resynthesis error term is not simple. System 3 (shown in cyan), for example, appears to enter into a regime of increased resynthesis error when $\left|\left|\mathcal{M}\right|\right|^2$ increases past 0.004, regardless of Q.

In general, however, as Q increases, resynthesis error decreases. One might presume this has to do with increased accuracy in eigendecomposition estimation, however, we can see a similar pattern from Test 2, which uses an exogenously supplied $\mathbf{\Lambda}$, c.f. figure 4.3.

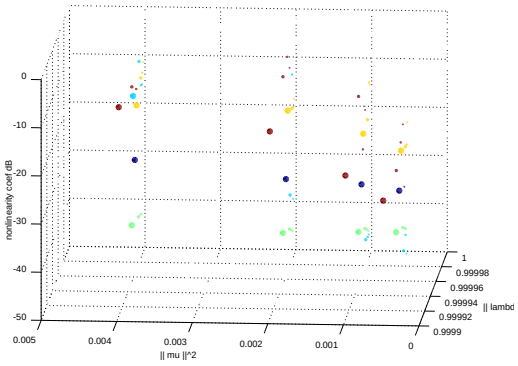

Figure 4.1: Battery 0, Test 1. Y-axis shows Φ. Each color corresponds to a unique coupling topology under study. While each system's response to an increase in $\left|\left|\mathcal{M}\right|\right|^2$ seems to be distinct, there is in general a trend correlating $\left|\left|\mathcal{M}\right|\right|^2$ with Φ for each system.

Table 4.1: Results for Battery 0. Mean and Standard Deviation calculated over all 7 systems.

$Qscale$	$\mathcal{M}$ scale	$E(\Phi_0)$	Std Φ_0	$E(e_{\exp 0})$	Std $e_{\exp 0}$	$E(\Phi_1)$	Std Φ_1	$E(e_{\exp 1})$	Std $e_{\exp 1}$	$E(\Phi_2)$	Std Φ_2	$E(e_{\exp 2})$	Std $e_{\exp 2}$
1	4	-35.7416778427	10.4955833369	-33.1765743073	2.3600440916	-31.6763009787	6.7584504951	-33.1769468884	2.3606614684	-31.6499411023	6.7788943568	-33.4454707303	2.8298067706
1	8	-34.6632988093	12.1193062877	-34.0204522277	2.6107699109	-30.226856977	8.2542576695	-34.0214965523	2.6112503565	-30.1612674491	8.2685424151	-34.6304376543	3.0820890397
1	16	-32.9659702758	13.5702337143	-31.0787241333	5.150229069	-28.7722248366	10.2666223706	-31.0821691996	5.1487821355	-28.0303510102	10.8335414678	-32.2450380404	3.2868933094
1	32	-25.415385535	12.8455889692	-31.3726997233	6.6230921719	-23.3424508583	10.1540743073	-31.434945233	6.6431634478	-23.2128831867	9.8007979107	-31.2679109347	5.9477587381
2	4	-35.9418374171	17.1018913798	-35.0964951535	5.3654606114	-31.8273316377	13.9318439762	-35.0983078527	5.3621120053	-31.784862349	13.8905460871	-35.7030338124	4.4764249174
2	8	-29.5267826745	20.5078396648	-34.9919539261	4.4406434532	-27.9846321392	18.5898568434	-35.0035210615	4.4261301419	-27.7402461686	18.6914563721	-35.6870528814	5.2522244199
2	16	-26.0499176879	17.0123798438	-30.5903677815	6.4226322137	-25.6362974343	16.1349211119	-30.7521361128	6.5214952468	-24.3922027706	16.8362235649	-32.0255486723	6.0400387744
2	32	-18.7676717484	16.5846503429	-31.1167833033	8.5534999857	-18.2665955372	14.6704651271	-31.5220115818	8.6827354074	-16.179330591	15.5829177907	-28.8783475204	7.1587751354
3	4	-35.9514732421	11.7674311824	-37.5108538855	1.5195680272	-34.9081932	10.0761050596	-37.5113965044	1.5201089413	-34.8712809249	10.0706569062	-37.7563812458	1.855576104
3	8	-31.9904531126	18.1659041376	-34.8306237947	8.9496327343	-30.4477001682	16.0094726831	-34.8380995054	8.9408596274	-29.9963681063	16.4178511305	-36.3596002856	6.4966745331
3	16	-21.5453227547	17.1250544526	-28.2879803245	8.4798454084	-21.1708652827	15.3552308919	-28.3156136888	8.4785619662	-20.587815325	15.5007295549	-28.7372707423	7.869202307
3	32	-21.58533766	15.2127092243	-32.7346325768	6.1311070116	-21.0728773618	13.2963381725	-33.0575080609	5.903065017	-22.869957874	11.680525753	-31.5047956609	6.3275633075
4	4	-31.8954585407	21.5125273083	-35.3083113751	7.6822107651	-30.3880972991	19.461771183	-35.3124082425	7.6773449418	-30.2987141604	19.4208433936	-36.0866828067	8.4359278353
4	8	-29.6535644607	21.2969789157	-32.058036227	9.9806886684	-28.0323980052	19.2197796977	-32.0675101012	9.975410261	-27.3437134961	19.7200232215	-32.7469748375	8.7120554982
4	16	-28.6449532171	18.156248979	-34.8367724412	5.0002901943	-27.3274441366	16.4993939231	-34.8564034372	4.9870686412	-26.5343381904	17.6831344901	-33.8931469752	5.9932427942
4	32	-22.7077307734	19.0774702533	-31.9480275521	8.9080930298	-22.0305151611	16.7386138111	-32.8402050649	9.1119954516	-18.081262951	20.5137633227	-27.7983000406	8.4582373957

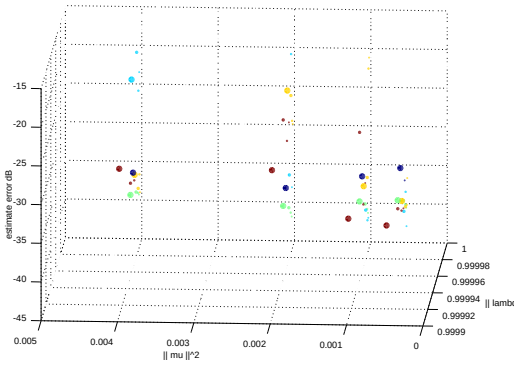

Figure 4.2: Battery 0, Test 1. Y-axis shows e_{exp}. Each coupling topology is associated with a different color. Notice how the individual system trends are more complex. Generally speaking, the variance of e_{exp} is smaller than that of Φ .

4.1.2 Battery 1: Error as a Function of Sampling Rate

The analysis techniques are effected by an increase in sampling frequency in a number of ways which can be difficult to disentangle. Since the amount of information in the signal is fixed as soon as the synthesis is run, a simple upsampling procedure is an insufficient method for simulating this.

One particularly meaningful and controllable side-effect of upsampling is the slowing of the linear system dynamics. This is quantitatively recreated by adjusting $\mathbf{\Lambda}$ such that the associated time-constants are increased as a linear function of the upsampling ratio. For a doubling of the sampling frequency, $\mathbf{\Lambda}$ effectively becomes $\mathbf{\Lambda}^{\frac{1}{2}}$. This is generally a welcome effect, as a slower system permits more measurements to be made

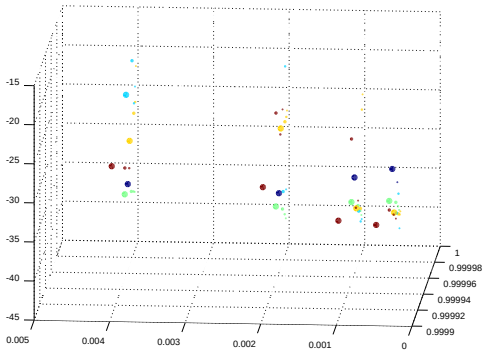

Figure 4.3: Battery 0, Test 2. Y-axis shows e_{exp}, for exogenously supplied eigenbasis. The pattern is similar to figure 4.2

per unit time, and thus a clearer picture of the underlying dynamics may generally result.

The rate at which the system is sampled also has an impact on our measurement of the noise term $\boldsymbol{\xi}$. This effect is due to the commonly observed $\frac{1}{f}$ spectral shaping of many forms of noise. Its appearance is hypothesized to have a negative impact on the ability to discern a valid eigendecomposition, since such a curvature is likely to be interpreted as a single real pole at DC. Furthermore, as this effect serves to dampen higher frequencies in the signal, the activity of the system in those higher frequency ranges will be harder to classify.

We begin by testing both of these effects on the accuracy of parameter estimation, while simultaneously observing secondary effects such as convergence. In this suite, 40 models in total were studied, each of order $N = 4$, with 20 allocated for the first battery and 20 for the second. Since this suite required the variation of intrinsic parameters,

each trial required a complete resynthesis of signal data. For both batteries, 3 trials were conducted on each model, wherein the requisite parameter was adjusted by factors of two. For each trial, the standard 3 tests were performed, as per the introductory notes of this chapter. For all tests, the number of microphones was presumed to equal the linear dimensionality of the system dynamics.

4.1.2.1 Battery 1, Suite 0: Slowing the Linear Dynamics in Λ

4.1.2.1.1 Methods

To slow down the linear component of the model dynamics, we note that $\tilde{\mathbf{\Lambda}}_a = \mathbf{\Lambda}^{\frac{1}{a}}$, where $\tilde{\mathbf{\Lambda}}_a$ is the identical linear transformation to $\mathbf{\Lambda}$ with a time constant scaled by a. This test battery subjected 20 different models to two trials and a control, examining the effects of doubling and quadrupling the system time constant. The error term e_{exp} and the nonlinearity index Φ are calculated for each of the three tests defined above in section 4.1.

4.1.2.1.2 Results

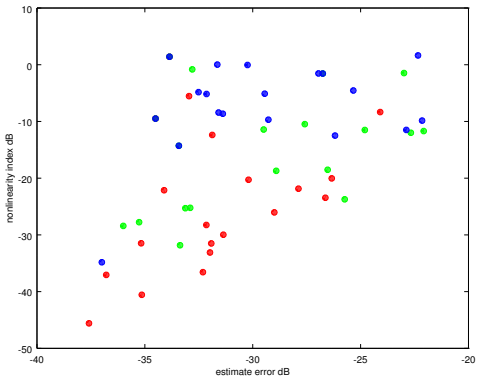

Figure 4.4: Battery 1, Suite 0: Scatterplot showing resynthesis error versus nonlinearity index. Red: Trial 0 ($\mathbf{\Lambda}$), Green: Trial 1 ($\mathbf{\Lambda}^{\frac{1}{2}}$), Blue: Trial 2 ($\mathbf{\Lambda}^{\frac{1}{4}}$)

Table 4.2: A Table containing the results of Battery 1, Suite 0 . Trial 0 uses control $\mathbf{\Lambda}$, Trial 1 uses $\mathbf{\Lambda}^{\frac{1}{2}}$, Trial 2 uses $\mathbf{\Lambda}^{\frac{1}{4}}$. Test 0 uses exogenous $\boldsymbol{V}$, Test 1 is pure estimates, Test No. 2 uses exogenous $\mathbf{\Lambda}$

Trial No.	Test No.	$\tilde{\mathbf{\Lambda}}_a$	Mean Φ (dB)	Std Φ (dB)	Mean e_{exp} (dB)	Std e_{exp} (dB)
0	0	$\mathbf{\Lambda}$	-25.0883	12.9993	-31.074	4.6930
1	0	$\mathbf{\Lambda}^{\frac{1}{2}}$	-14.9944	10.6851	-28.536	5.7966
2	0	$\mathbf{\Lambda}^{\frac{1}{4}}$	-7.3818	8.5215	-28.129	5.6321
0	1	$\mathbf{\Lambda}$	-24.1092	12.4313	-31.793	3.5652
1	1	$\mathbf{\Lambda}^{\frac{1}{2}}$	-14.8823	10.1898	-29.621	4.6303
2	1	$\mathbf{\Lambda}^{\frac{1}{4}}$	-7.3103	8.2407	-29.454	4.3216
0	2	$\mathbf{\Lambda}$	-23.5788	12.2210	-31.974	3.3465
1	2	$\mathbf{\Lambda}^{\frac{1}{2}}$	-12.2585	9.3199	-16.658	4.2177
2	2	$\mathbf{\Lambda}^{\frac{1}{4}}$	-7.2653	8.5285	-12.685	3.9706

4.1.2.2 Battery 1, Suite 1: Reducing High-Frequency Information in $\boldsymbol{\xi}$

4.1.2.2.1 Methods

To simulate the spectral curvature in common forms of noise, a single real pole was placed on $\boldsymbol{\xi}$ prior to applying the nonlinear filterbank. The center frequency of this pole was varied each of the two trials. To isolate the effects of the 1-pole filter from any overall gain change in $\boldsymbol{\xi}$, make-up gain was applied post filtering. Although this pre-processing filter will attenuate a portion of the high-frequency dynamics of the system in general, it cannot be commuted with the nonlinear contribution to dynamics. Thus, this pre-processing only simulates the action of upsampling in the presence of bandlimited noise, and not the presence of roll-off in the measurements.

4.1.2.2.2 Results

The mean estimated resynthesis error for the two trials does not diverge significantly from the control. There appears to be an effect on the variance of this metric, however. The resynthesis errors in Trial 1, with $f_c = \pi$ appear to vary much less than

Table 4.3: A Table containing the results of Battery 1, Suite 1 . Trial No. 0 is unfiltered, Trial No. 1 $f_c = \pi$, Trial No. 2 $f_c = \frac{\pi}{2}$. Test 0 uses exogenous $\boldsymbol{V}$, Test 1 is pure estimates, Test No. 2 uses exogenous $\boldsymbol{\Lambda}$

Trial No.	Test No.	f_c	Mean Φ (dB)	Std Φ (dB)	Mean e_{exp} (dB)	Std e_{exp} (dB)
0	0	N/A	-29.600	13.988	-30.819	6.6419
1	0	π	-31.116	12.949	-32.299	4.9423
2	0	$\frac{\pi}{2}$	-28.154	14.654	-31.402	6.8217
0	1	N/A	-29.322	13.500	-31.121	6.1208
1	1	π	-30.479	12.832	-32.828	4.5717
2	1	$\frac{\pi}{2}$	-28.060	13.761	-31.790	6.3179
0	2	N/A	-28.937	13.742	-32.414	5.1248
1	2	π	-30.231	12.796	-33.451	4.4199
2	2	$\frac{\pi}{2}$	-27.924	13.699	-32.501	6.3220

those of Trial 2, with $f_c = \frac{\pi}{2}$. We look for any correlation between Φ and resynthesis error by plotting the two parameters against each other in a scatterplot.

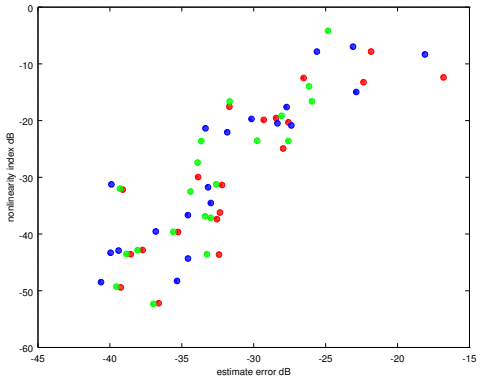

Figure 4.5: Battery 1, Suite 1. Scatterplot showing resynthesis error versus nonlinearity index. Red: Control (no filter on $\boldsymbol{\xi}$), Green: Trial 1 ($\boldsymbol{\xi}$ filter $f_c = \pi$), Blue: Trial 2 ($\boldsymbol{\xi}$ filter $f_c = \frac{\pi}{2}$)

4.1.3 Battery 2, Suite 0: In-Sample Error and Improvement as a Function of Analysis Window

4.1.3.1 Methods

It is intuitive that the analysis algorithms presented might benefit from as much data as can be supplied. Therefore, it would seem that the length of the timeseries analyzed should be as long as possible. However, the increase in complexity with respect to the timeseries provides us with an incentive to limit its length. That said, it is unclear the extent to which the accuracy will benefit from increased length versus an increased number of simultaneous measurements. Therefore, it is necessary to conduct a test battery wherein the accuracy as a function of window size can be determined.

To this end, 10 separate systems were synthesized, upon which 10 analysis trials were performed. Each trial used a different number of timesteps to conduct its analysis, comprised of the now-familiar 3 individual variants of the analysis algorithm. The possible window sizes were 98, 195, 391, 781, 1563, 3125, 6250, 12500, 25000, and 50000. For Suite 0, analysis was performed over the entire window and residuals were calculated on the same data. Therefore, Suite 0 measures in-sample error.

4.1.3.2 Results

The results appear to indicate that the peak expected in-sample performance of the algorithm and the peak expected in-sample contribution of the nonlinear estimation step to that accuracy do not occur at the same window size. The trend for window sizes below 391 samples tells a different story. Recall that Suite 0 measured in-sample residuals. These low-error levels are likely the result of overfit.

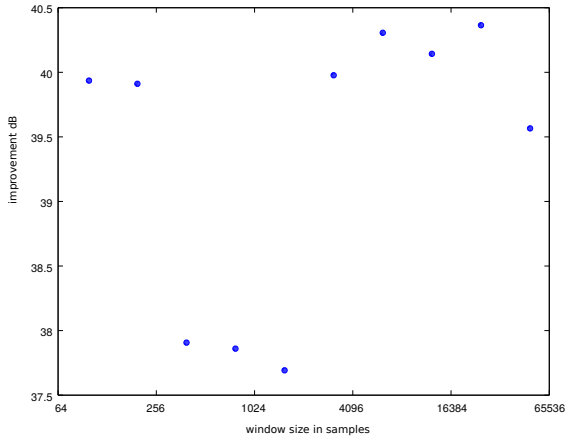

Figure 4.6: Battery 2, Suite 0: In-Sample $E\left(e_{\Delta}\right)$ as a Function of window size. The region between Trial 5 and 8 (3125 - 25000 samples) appears to maximize the improvement best.

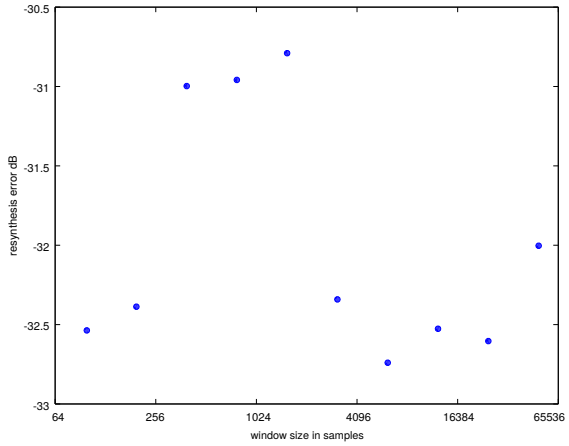

Figure 4.7: Battery 2, Suite 0: In-Sample Resynthesis Error $E\left(e_{\text{exp}}\right)$ as a Function of window size.

4.1.4 Battery 2, Suite 1: Out-of-Sample Error and Improvement as a Function of Analysis Window

4.1.4.1 Methods

To take out-of-sample measures, we rely on the ground truth recording of state and total RMS of the signal over a much longer domain– 100000 samples– than was used in the Trials in Suite 0. Since we are using the ground truth state vector sequence, we

Table 4.4: Results for Battery 2, Suite 0: Mean and Standard Deviation calculated over all 10 systems. Notice the spike in $E\left(e_{\exp}\right)$ between Trial 1 - Trial 3. Also, recall that this Suite measures in-sample performance, which is not strictly an indicator of future performance in general.

Window Size	$E\left(\Phi_0\right)$	Std Φ_0	$E\left(e_{cv0}\right)$	Std e_{cv0}	$E\left(e_{\Delta}\right)_0$	Std $e_{\Delta 0}$	$E\left(\Phi_1\right)$	Std Φ_1	$E\left(e_{cv1}\right)$	Std e_{cv1}	$E\left(e_{\Delta}\right)_0$	Std $e_{\Delta 0}$	$E\left(\Phi_2\right)$	Std Φ_2	$E\left(e_{cv2}\right)$	Std e_{cv2}
196	-28.1807467484	8.9502515488	-31.2626276701	4.7954928545	38.7866435841	12.0170336827	-26.5060057029	8.7386121002	-32.3871790435	3.65765104	39.9111949575	11.1178320046	-23.3954203741	11.8616009994	-31.4990723354	6.0095230831
392	-29.6856285332	13.5100973751	-30.1258854798	5.8570657039	37.0354299157	12.9209333627	-25.9035601216	12.7138989807	-30.9973120495	5.3296167266	37.9068564854	12.3863918882	-25.3596401927	12.7201124056	-31.7366840594	5.4811845389
782	-31.261378284	14.6427126305	-30.2125225942	5.8488006623	37.1141802728	12.7605708799	-26.6785050074	13.3108990836	-30.9585723799	5.4143916878	37.8602300585	12.2870803632	-26.1114472386	13.2658035812	-31.7615340535	5.5482230854
1564	-32.0907201792	16.0306316588	-30.1860061592	5.8868768799	37.0882042818	12.8897780218	-26.5807230343	13.6922806001	-30.7900915873	5.5074017117	37.6922897099	12.5124225092	-26.0223235706	13.6143472689	-31.7340648966	5.5628673126
3126	-34.4261549333	13.4941140973	-31.6101298786	5.2146125782	39.2460824125	12.6949976219	-28.4344101813	11.1704814265	-32.3412421185	4.4793071597	39.9771946524	12.1159336666	-25.4536500645	12.9011177397	-32.2036719754	5.6053753258
6251	-34.2620693251	13.5824711313	-32.0993371207	5.176078212	39.6652098273	12.6227488312	-28.1492986193	11.0001036983	-32.7401974642	4.4018915613	40.3060701708	12.0721703594	-25.2719604339	12.5341316787	-32.4458937065	5.8128997932
12501	-34.3972298927	13.2080351433	-32.0364615722	5.2150466681	39.6534430714	12.689526364	-28.1327859913	10.6498617793	-32.5262411361	4.5233168108	40.1432226352	12.2238742418	-25.3514909298	12.0338835149	-32.1866830435	5.9436092129
25001	-34.6153229032	12.4008939352	-32.1650857688	4.911806935	39.9258049645	12.3067178036	-28.1033469433	9.8203259145	-32.6038683499	4.277356563	40.3645875457	11.8999683154	-24.8670248295	12.2948750192	-30.440591607	10.370187384
50001	-33.3491200109	12.5683560183	-31.4716686956	5.0907636891	39.0342002589	12.0727417647	-26.62751244	10.2158448598	-32.0032472541	4.6137463884	39.5657788174	11.7040394192	-24.5593647141	10.8485960255	-32.0172394662	5.7258507571

can derive a sample-by-sample error analogous to the one used in Battery 2, Suite 0, but measuring the system's performance on new examples of the same system.

4.1.4.2 Results

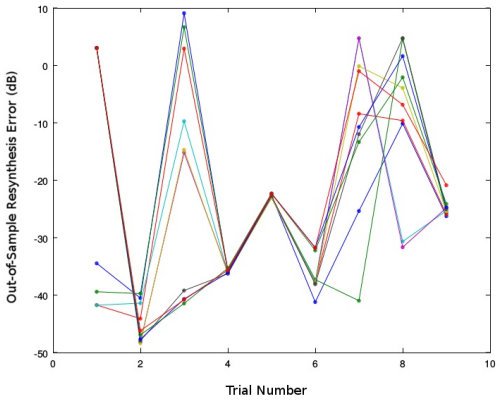

Figure 4.8: Battery 2, Suite 1: Out-of-Sample $E\left(e_{\text{exp}}\right)$ as a Function of window size. Note the optimal region is different from 4.7.

The results clearly show an optimal region for out-of-sample resynthesis acuity between 781 - 3125 samples. In this region, not only has the expected resynthesis error been minimized, but furthermore the variance appears to be well under control. This suggests our results will be more consistent in this range as well.

4.1.5 Error as a Function of Channels

4.1.5.1 Methods

A battery of tests was designed to determine the effect of the number of microphones on the analysis. The synthesized signals were generated using $N = 6$ so that the following microphone configurations could be used: 1, 2, 3, and 6. Microphone configurations with 4 or 5 channels were avoided at this time for simplicity. The arrangement of the microphones, encoded by $\boldsymbol{W}$, was held fixed across these trials, with each trial

virtually muting or unmuting a subset of the microphone recordings. This assuages the confounding effects of $\boldsymbol{W}$ on the analysis.

4.1.5.2 Results

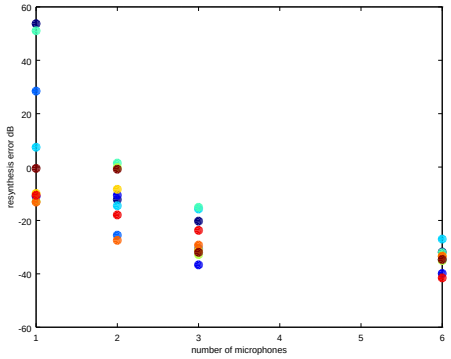

Figure 4.9: Battery 3: $E\left(e_{\text{exp}}\right)$ as a function of microphone quantity. Note the diminishing returns as microphones are added. Each color corresponds to a particular system.

The results show a diminishing improvement in accuracy as the number of mics is increased to N. There also appears to be a decrease in standard deviation, which suggests more reliable analyses.

4.2 Retrieval of Parameters from Field Measurements

4.2.1 Methods

Two batteries of tests were applied to the white noise recordings described in section 1.1.1.2.2. In the first battery, we applied the LPC-based linear analysis, whereas in the second, we supplied an *a priori* estimate of $\mathbf{\Lambda}$ and used the techniques described in section 3.4. Each trial in these batteries comprised of a single test, wherein the nonlinear parameters were estimated using the algorithm. It is hypothesized that this analysis technique is sensitive to the order of the analysis, the length of the window used, and the offset into the data. Thus, the first battery varied each of these parameters to find an optimal combination of parameters. Since $\hat{\mathbf{\Lambda}}$ was supplied to the analysis routine in the second battery, only the length and offset of the analysis window were varied.

4.2.2 Results

The results for Battery 1 are shown in figure 4.10. This demonstrates, among other facts, that the LPC-based linear analysis is fairly unstable. The number of valid data points, including even suboptimal ones, is much smaller than the total number of trials performed. Unfortunately, applying LPC to highly nonlinear data occasionally results in the placement of a pair of poles on the real axis, which results in a singular eigenbasis. Only the analyses based on valid eigenbases are shown.

Furthermore, in figure 4.10, we can see that many analyses, though valid, are suboptimal, because either e_{exp} is positive, or e_{Δ} is negative. If e_{exp} is positive, then the total variance explained by the nonlinear estimate is greater than the total recorded variance of the system, which obviously suggests the nonlinear estimate is incorrect. If e_{Δ} is negative, then the linear component of the analysis explains the system dynamics more optimally than the nonlinear component. Both of these measures could be used as a validation

method for the analysis parameters, and placement into any except for the lower-right quadrant of figure 4.10 indicates a poor selection of those parameters.

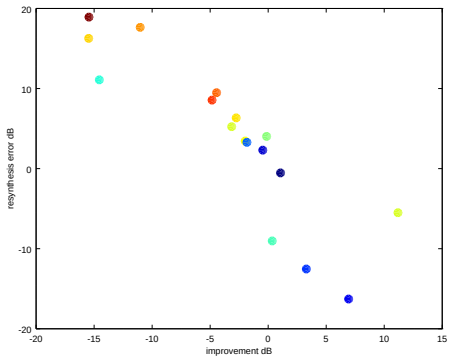

Figure 4.10: A scatterplot of e_{Δ} vs e_{exp} for LPC-based analysis (Battery 1). The optimal region is the lower right quadrant. Colors (in rainbow order) represent Φ.

The results of the Battery 2 are showin in figure 4.11. As is evidenced from the much more crowded scatterplot, the analysis is valid for all datapoints because the eigenbasis is given and invertible. However, unfortunately, the number of optimal analyses has not changed much. This suggests that one or more parameters have still been poorly selected.

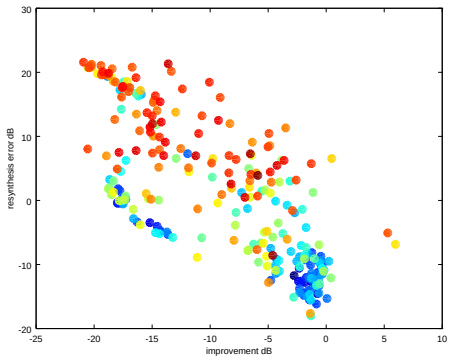

Figure 4.11: A scatterplot of e_Δ vs e_{exp} for exogenous-$\hat{\mathbf{\Lambda}}$ analysis (Battery 2). The optimal region is the lower right quadrant. Colors (in rainbow order) represent Φ.

Chapter 5

Conclusion

5.1 Research Questions

In the preceding chapters, an algorithm for analysis and synthesis of nonlinear systems has been derived and validated. We now present a brief account of the assessment, an anecdote regarding the genesis of this project, and an overview of possible future research directions.

5.1.1 Bandwidth of Measurements

How does sampling rate affect our ability to estimate the model dynamics? In the range of parameters studied, slowing the linear dynamics seems to reduce e_{exp}, but these returns diminish substantially with subsequent upsampling. The effect of expanding the measured bandwidth further beyond that of the excitatory noise ξ appears only to impact our estimate of Φ in any meaningful sense, leaving e_{exp} unchanged.

5.1.2 Window Size

How does the covariance window size affect our ability to estimate the model dynamics? The results for mean in-sample resynthesis error suggest that the overall peak improvement in expected accuracy as a function of window size was observed at 62500 time-steps. However, the results for mean in-sample nonlinear improvement suggest the benefits of using the nonlinear parameter estimation steps are most significant at or just above this size. This points to an optimal range of window sizes between 4000 and 10000.

For out-of-sample error, we observed an optimal range between 781 - 3125 samples. Although trial 9 shows a modest improvement over trial 8, we can disregard it as an option because of the much more consistently accurate out-of-sample measurements which are available at far lower computational cost. Recall that each trial increment represents a doubling of term data to be processed.

These results actually suggest that, for previous batteries, overall accuracy in determining both linear and nonlinear contributions to system dynamics might be improved by reducing the window size. Furthermore, these findings suggest that the linear and nonlinear portions of the solution might benefit from analyzing different timescales.

5.1.3 Dimensionality of Measurements

How does the measurement dimensionality affect our ability to estimate the model dynamics? The results suggest an overall improvement may be expected as channels are added, however the greatest improvement was observed between 1 and 2 dimensions, with increasingly modest improvements in both accuracy and reliability of estimates with subsequent increases in dimensionality.

5.1.4 Parameter Estimation of Recordings

One way to assess the accuracy of a given analysis is to generate synthetic model output and compare it with the recorded data. This comparison could comprise a number of different tests, but since ostensibly our goal is to generate computer music, perhaps the most relevant test is to listen to the results directly. We selected 4 estimates from each battery which are within the optimal region, and resynthesized the model dynamics using the estimated parameters. For the purposes of this document, we include spectrograms of the recordings as well as links to the audio.

A disclaimer: they are not pretty. The precise reason for this lack of musicality is unclear at the time of this writing. However, the analysis signal from which the models were derived appears to suffer from the same lack of aesthetic beauty. Perhaps the linear components of the system are simply too damped, or the system nonlinearities are too strong.

Although these results may not be the most musical examples of the synthesis technique, they do offer us a glimmer of possibility for future explorations in analysis/resynthesis. At the time of this writing, the most compelling examples of the synthesis technique apply parameter sets derived from user interaction. A few of these will also be included as spectrograms and recordings.

5.2 Future Work

This dissertation has produced more questions than answers. Future work will attempt to address some of these new questions, most likely after a re-implementation of the algorithms using C++. The affordances of modern C++ will result in a flexible, efficient and maintainable library, well-suited to real-time processing in an environment like PureData, SuperCollider, or as an audio plugin. Current prototypes made use of

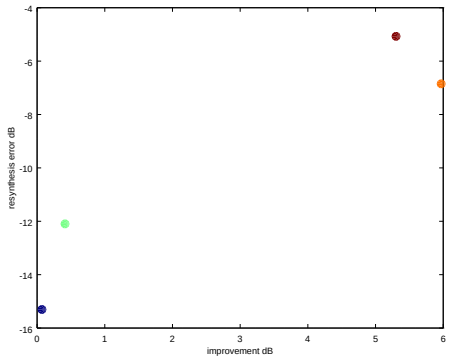

Figure 5.1: A scatterplot of e_{Δ} vs e_{exp} for exogenous-$\hat{\mathbf{\Lambda}}$ analysis (Battery 2). This plot shows the datapoints selected because they lie in the optimal region. Colors (in rainbow order) represent Φ.

the Octave scientific computing platform, and as a result these prototypes are slow and inappropriate for musical exploration. The C++ implementation process will begin with the synthesis technique so that new parameter ranges and control paradigms can be expermimented with.

Future work will also include attempting to integrate the analysis technique with more robust linear estimation techniques, such as [Smi07]. Recall that we defined in section 3.4 a method for estimating the nonlinear least-squares parameters given exogenously provided $\hat{\mathbf{\Lambda}}$. Assuming we can determine the best basis for decomposing the measurements into state-space, we can apply the technique as presented in this section in tandem with any linear estimation technique. Of particular interest are techniques such as [LS86][p.32], which are capable of tracking time-varying estimates of $\hat{\mathbf{\Lambda}}$ and associated state-variables. This family of techniques operate by modeling $\hat{\mathbf{\Lambda}}$ as a random variable,

and encompasses the Kalman filter as a special case.

Still more questions arise when attempting to place the linear and quadratic problems into a single least-squares framework. So far, these questions remain unanswered, and this is no major impediment, since the current method appears to give satisfactory results while keeping these estimation problems separate. However, if it were possible to combine the two, the estimated $\hat{\mathbf{\Lambda}}$ parameters and associated terms such as $\overset{\boldsymbol{E}}{\boldsymbol{SEV}}_{\boldsymbol{x}}$, which are, ultimately, estimates themselves, could be updated given this step in the algorithm and the results fed back into the linear estimation steps. This feedback might allow us to recurse toward a better estimate of $\hat{\mathbf{\Lambda}}$ as well as a better estimate of $\hat{\boldsymbol{V}}$. The difficulty of this approach comes from the dimension reduction step described in section 3.3.3.

Another line of inquiry relates back to the instrumentation devised to make the recordings and performances described in section 1.1.1. Rather than simply increasing the number of scalar field measurements, the advantage concomitant with increasing the dimensionality of the measurements might also be achieved by increasing the dimensions of each measurement. A simple way to achieve this involves using PIN photodiodes with two axes of photoelectric response. These measurements might simply be treated as unstructured additional coefficients in our calculations, or they might be used to derive entirely new algorithms which operate on sets of complex measurements. Although preserving this structure will pose still new challenges, some authors (C.F. [MG09, p. 13]) suggest that such techniques handily improve performance metrics in real-world data.

5.3 Inspiration

The idea for this algorithm came from an admittedly unlikely place: watching a spider build a web. Paralyzed with arachnophobia, I observed the structural threads of the

web vibrating pleasantly both as a result of the spider as she worked and the wind lightly bowing laterally across their tiny surfaces. I recalled from a biology class that most species of spiders actually cannot see particularly well, and I pondered how odd it was that such an animal, a high opportunistic predator in its niche, could come to have such a shortcoming. I imagined that these spiders must have a particularly well-developed listening ability, not for vibrations in air, but those in their webs. These networks of sheer fibers must act as their prosthetic ears.

Naturally, I wondered how I could tap into the sounds that the spiders could hear. At that point, a large stratum of my creative output revolved around amplifying solid objects using piezoelectric contact microphones and hydrophones. One basic rule-of-thumb for these techniques is that the structural wave medium to be measured should be at least as massive as the microphone itself. As such, I had spent a good deal of my time working with instruments made of materials like metal, wood, and bone. This new line of inquiry would clearly require the development of some new tools. I began to do some cursory research into arachnology, firstly to validate my hypothesis about the ecological need for spiders to hear through their webs, and if so, to determine what sort of microphone I would need to purchase or construct.

As it would happen, my hypothesis was essentially substantiated within the literature. Furthermore, several papers described the variations in spider auditory processing systems- located primarily at the base of each leg– which seemed to be correlated with the variations in web architecture. For example, spider species who build disorderly "cobwebs"- including the infamous black widow, among others- generally had better temporal acuity than their cousins who build iconic "orb-webs", who instead appear quite adept at distinguishing between adjacent frequencies. Whereas inter-leg time differences might provide more useful information for an erratic web with more damped modes, more orderly webs with more underdamped modes might suggest a frequency-domain

analysis on the part of the animal.

Numerous measurement methods were proposed in the literature. Many of them were impractical for my particular use-case. Several papers made mention of optical techniques as an in-vivo method for measurement. Simultaneously, I began to think of other possible amplification subjects uniquely suited to optical microphones. Many of them shared the property that the implements of piezoelectric amplification would overwhelm their dynamics. I thought mainly of extremely light objects or liquids. While only a single spatial point of measurement was typically discussed in the entomology literature, I began to dream of a vast network of microphones, recording the smallest deviations in displacement, over a wide bandwidth. I also began to dream of simulating and analyzing those deviations, perhaps as a way to see into the underlying dynamics which had previously been hidden.

This inspiration sparked a long journey that has encompassed research and experiments in optics, detours into both analog instrumentation and embedded systems, and numerous forays into linear algebra and digital signal processing. This document serves as a field guide to my experiences in some of these topics. That I can reach this journey's end with more questions than I had at its inception is testament to the fact that journeys themselves do not neatly begin and end. These paths of inquiry merely change into new ones, each seemingly more infuriating and fascinating than the next, uncountable transitions which hide this condition's underlying salience. And yes, I'm still terrified of spiders.

Appendix A

Final notes

A.1 In Tableau

$$
\Pi = \left[\begin{array}{cccccccc}
-\Pi(2,1;1)_{111} & \dots & -\Pi(2,1;1)_{N11} & \dots & -\Pi(2,1;1)_{NN1} & \dots & -\Pi(2,1;1)_{NNN} \\
\Pi(1,2;1)_{111} & \dots & \Pi(1,2;1)_{N11} & \dots & \Pi(1,2;1)_{NN1} & \dots & \Pi(1,2;1)_{NNN} \\
-\Pi(3,1;1)_{111} & \dots & -\Pi(3,1;1)_{N11} & \dots & -\Pi(3,1;1)_{NN1} & \dots & -\Pi(3,1;1)_{NNN} \\
\Pi(1,3;1)_{111} & \dots & \Pi(1,3;1)_{N11} & \dots & \Pi(1,3;1)_{NN1} & \dots & \Pi(1,3;1)_{NNN} \\
\vdots & & \vdots & & \vdots & & \vdots \\
-\Pi(N,1;1)_{111} & \dots & -\Pi(N,1;1)_{N11} & \dots & -\Pi(N,1;1)_{NN1} & \dots & -\Pi(N,1;1)_{NNN} \\
\Pi(1,N;1)_{111} & \dots & \Pi(1,N;1)_{N11} & \dots & \Pi(1N;1)_{NN1} & \dots & \Pi(1,N;1)_{NNN} \\
-\Pi(3,2;1)_{111} & \dots & -\Pi(3,2;1)_{N11} & \dots & -\Pi(3,2;1)_{NN1} & \dots & -\Pi(3,2;1)_{NNN} \\
\Pi(2,3;1)_{111} & \dots & \Pi(2,3;1)_{N11} & \dots & \Pi(2,3;1)_{NN1} & \dots & \Pi(2,3;1)_{NNN} \\
\vdots & & \vdots & & \vdots & & \vdots \\
-\Pi(N,2;1)_{111} & \dots & -\Pi(N,2;1)_{N11} & \dots & -\Pi(N,2;1)_{NN1} & \dots & -\Pi(N,2;1)_{NNN} \\
\Pi(2,N;1)_{111} & \dots & \Pi(2,N;1)_{N11} & \dots & \Pi(2,N;1)_{NN1} & \dots & \Pi(2,N;1)_{NNN} \\
\Pi(3,4;1)_{111} & \dots & \Pi(3,4;1)_{N11} & \dots & \Pi(3,4;1)_{NN1} & \dots & \Pi(3,4;1)_{NNN} \\
-\Pi(4,3;1)_{111} & \dots & -\Pi(4,3;1)_{N11} & \dots & -\Pi(4,3;1)_{NN1} & \dots & -\Pi(4,3;1)_{NNN} \\
\vdots & & \vdots & & \vdots & & \vdots \\
\vdots & & \vdots & & \vdots & & \vdots \\
-\Pi(N,N-1;1)_{111} & \dots & -\Pi(N,N-1;1)_{N11} & \dots & -\Pi(N,N-1;1)_{NN1} & \dots & -\Pi(N,N-1;1)_{NNN} \\
\Pi(N-1,N;1)_{111} & \dots & \Pi(N-1,N;1)_{N11} & \dots & \Pi(N-1,N;1)_{NN1} & \dots & \Pi(N-1,N;1)_{NNN} \\
\hline
\Pi(1,2,3;1)_{111} & & & \dots & & & \Pi(1,2,3;1)_{NNN} \\
(\Pi(3,1,2;1)-\Pi(2,3,1;1))_{111} & & & \dots & & & (\Pi(3,1,2;1)-\Pi(2,3,1;1))_{NNN} \\
\Pi(1,2,4;1)_{111} & & & \dots & & & \Pi(1,2,4;1)_{NNN} \\
(\Pi(4,1,2;1)-\Pi(2,4,1;1))_{111} & & & \dots & & & (\Pi(4,1,2;1)-\Pi(2,4,1;1))_{NNN} \\
\vdots & & & & & & \vdots \\
\Pi(1,2,N;1)_{111} & & & \dots & & & \Pi(1,2,N;1)_{NNN} \\
(\Pi(N,1,2;1)-\Pi(2,N,1;1))_{111} & & & \dots & & & (\Pi(N,1,2;1)-\Pi(2,N,1;1))_{NNN} \\
\Pi(1,3,4;1)_{111} & & & \dots & & & \Pi(1,3,4;1)_{NNN} \\
(\Pi(4,1,3;1)-\Pi(3,4,1;1))_{111} & & & \dots & & & (\Pi(4,1,3;1)-\Pi(3,4,1;1))_{NNN} \\
\vdots & & & & & & \vdots \\
\Pi(1,3,N;1)_{111} & & & \dots & & & \Pi(1,3,N;1)_{NNN} \\
(\Pi(N,1,3;1)-\Pi(3,N,1;1))_{111} & & & \dots & & & (\Pi(N,1,3;1)-\Pi(3,N,1;1))_{NNN} \\
\vdots & & & & & & \vdots \\
\Pi(1,N-1,N;1)_{111} & & & \dots & & & \Pi(1,N-1,N;1)_{NNN} \\
(\Pi(N,1,N-1;1)-\Pi(N-1,N,1;1))_{111} & & & \dots & & & (\Pi(N,1,N-1;1)-\Pi(N-1,N,1;1))_{NNN} \\
\vdots & & & & & & \vdots \\
\Pi(N-2,N-1,N;1)_{111} & & & \dots & & & \Pi(N-2,N-1,N;1)_{NNN} \\
(\Pi(N,N-2,N-1;1)-\Pi(N-1,N,N-2;1))_{111} & & & \dots & & & (\Pi(N,N-2,N-1;1)-\Pi(N-1,N,N-2;1))_{NNN}
\end{array}\right] \tag{A.1}
$$

A.2 Spectrograms

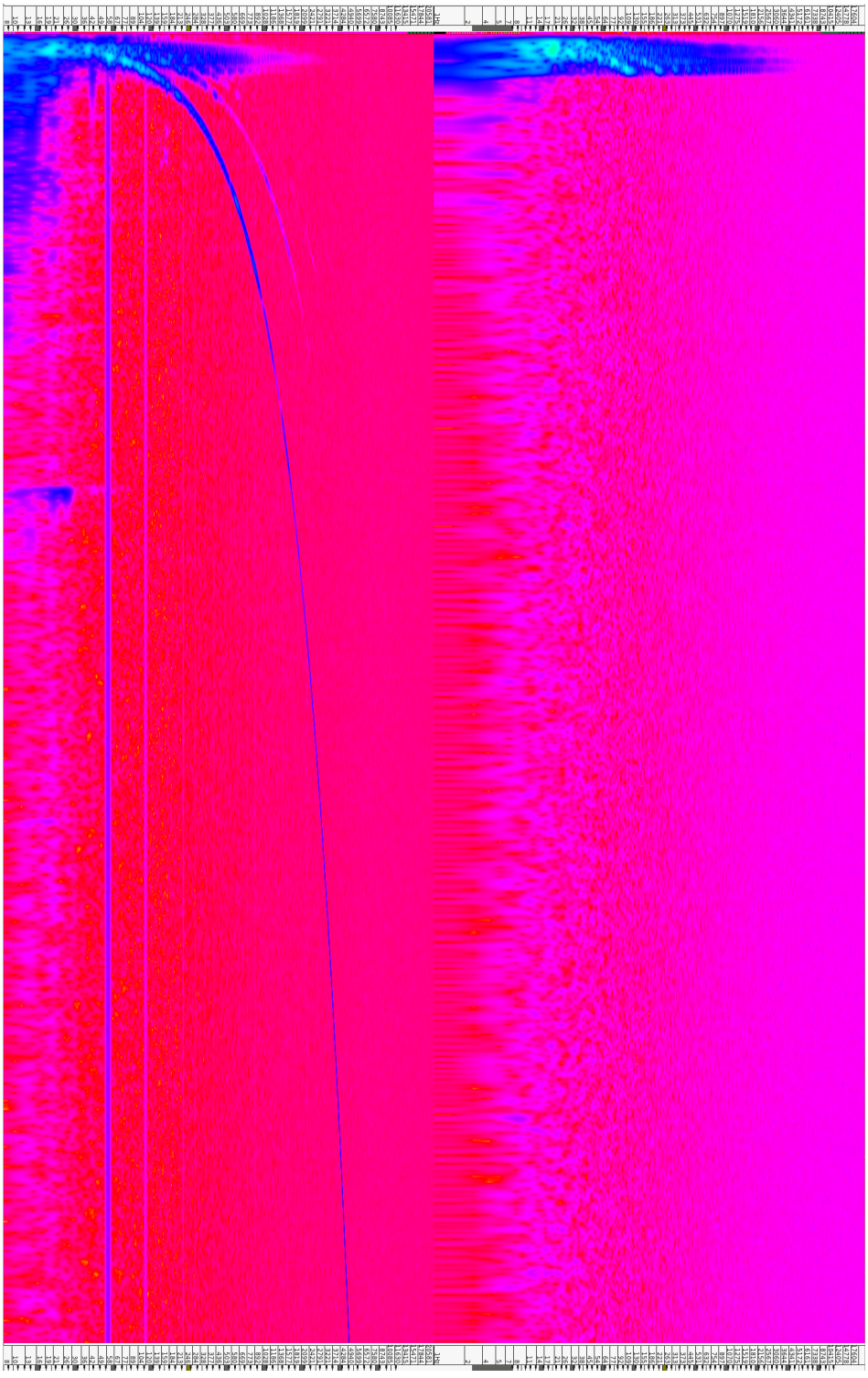

Figure A.1: Spectrogram of Test 00.00 (detail, showing 0→5 kHz). On the left is the signal from vibrometer "A", and on the right, the signal from vibrometer "B".

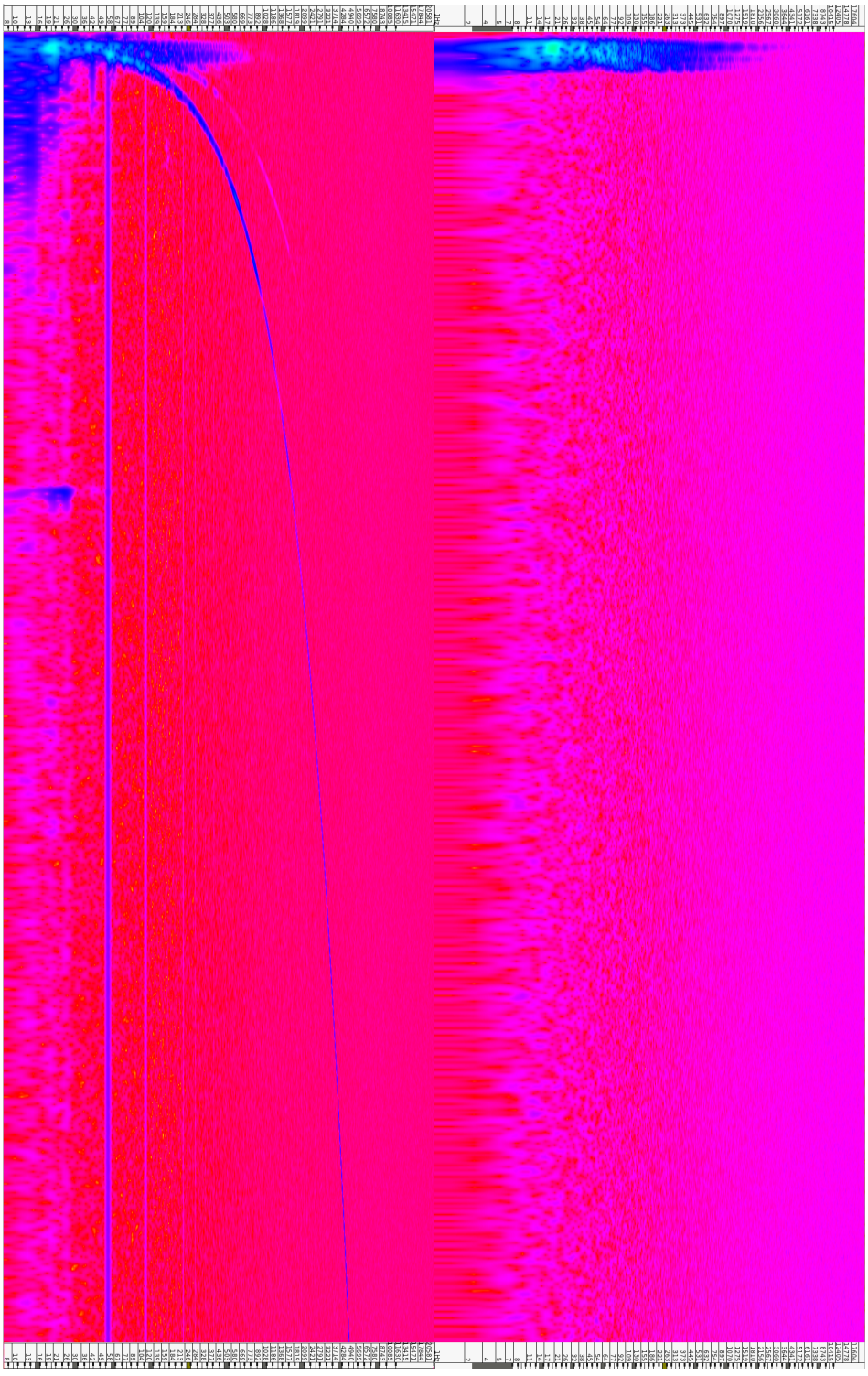

Figure A.2: Spectrogram of Test 00.01 (detail, showing 0→5 kHz). On the left is the signal from vibrometer "A", and on the right, the signal from vibrometer "B".

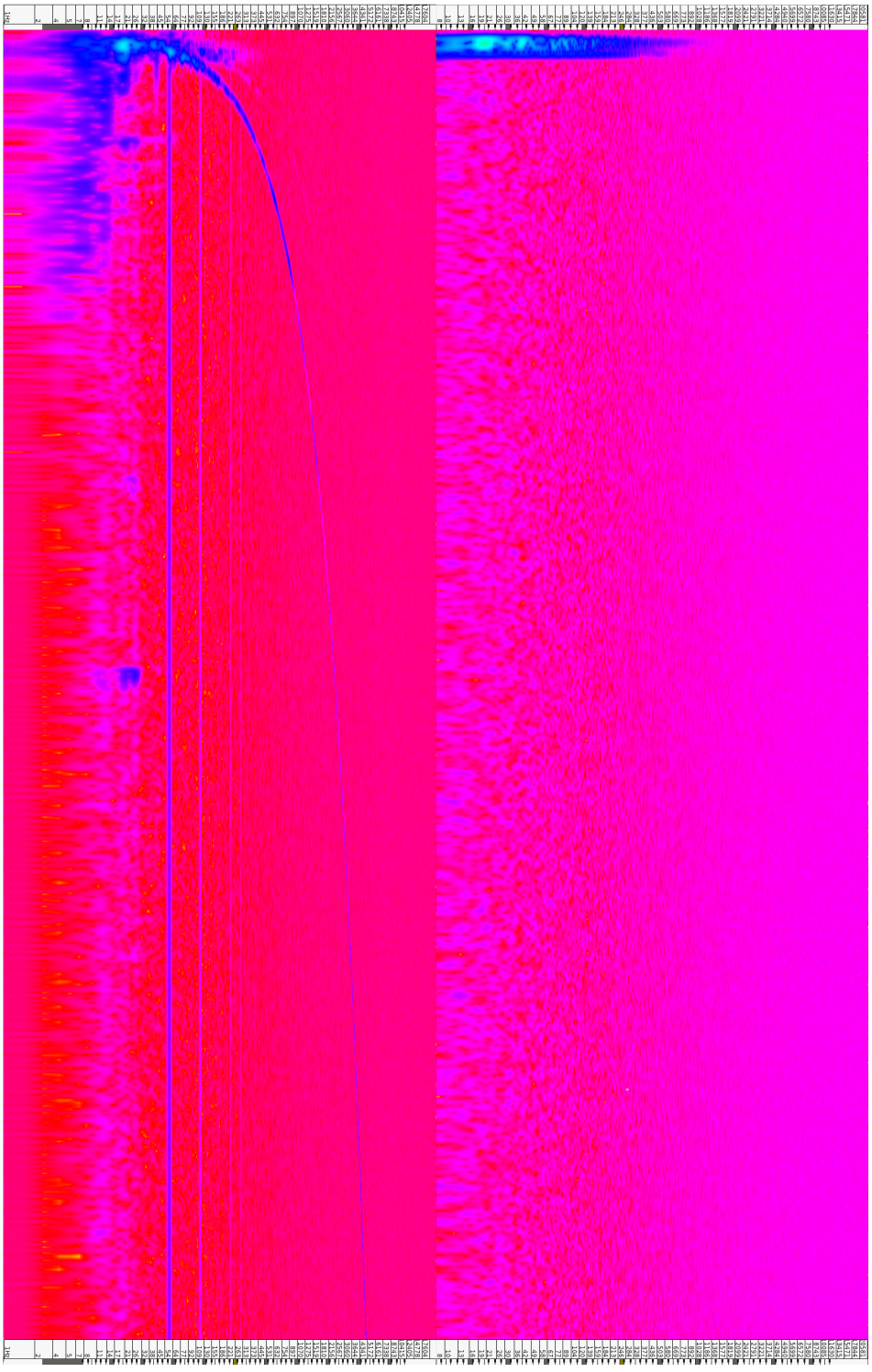

Figure A.3: Spectrogram of Test 00.02 (detail, showing 0→5 kHz). On the left is the signal from vibrometer "A", and on the right, the signal from vibrometer "B".

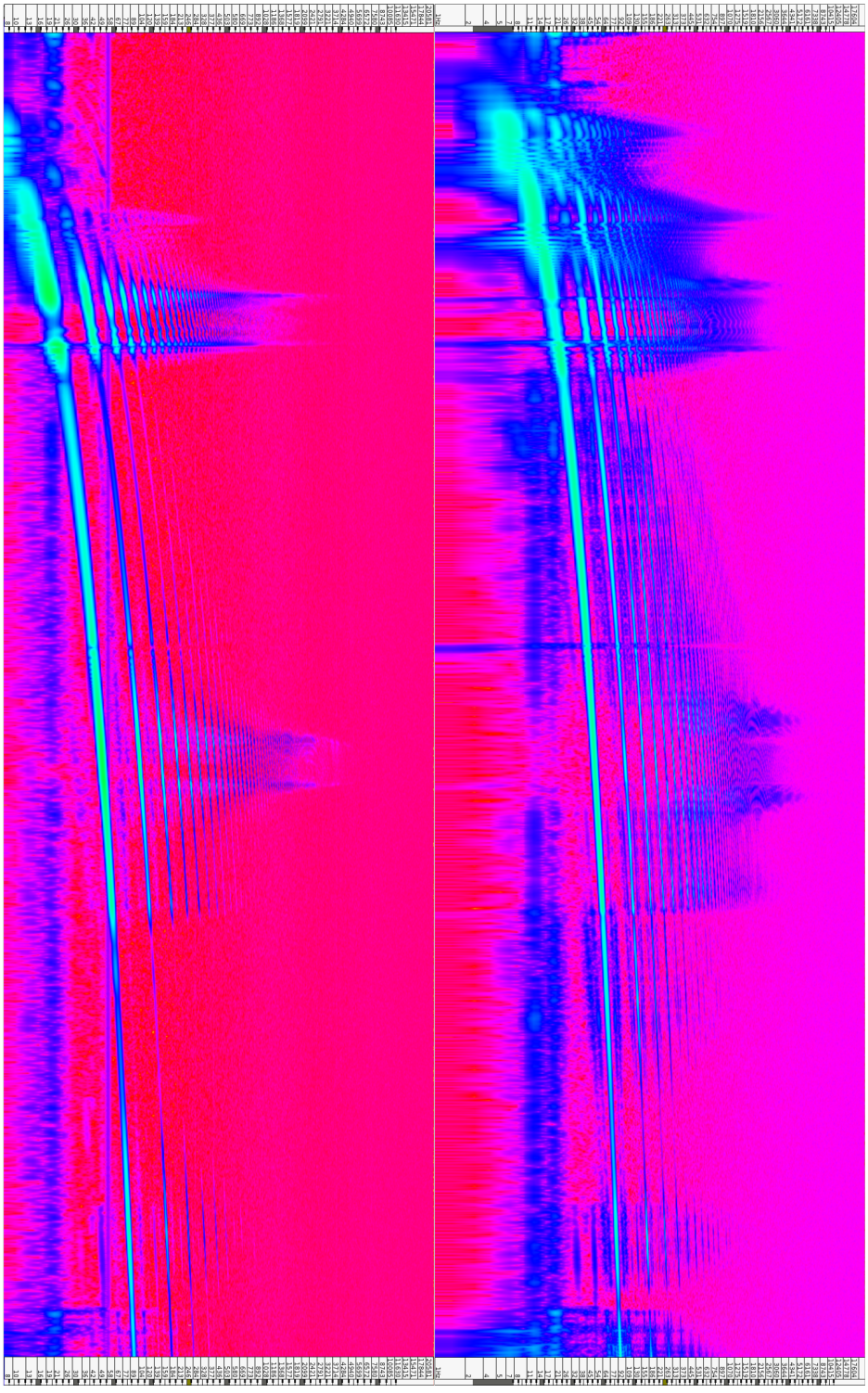

Figure A.4: Spectrogram of Test 00.03. On the left is the signal from vibrometer “A”, and on the right, the signal from vibrometer “B”.

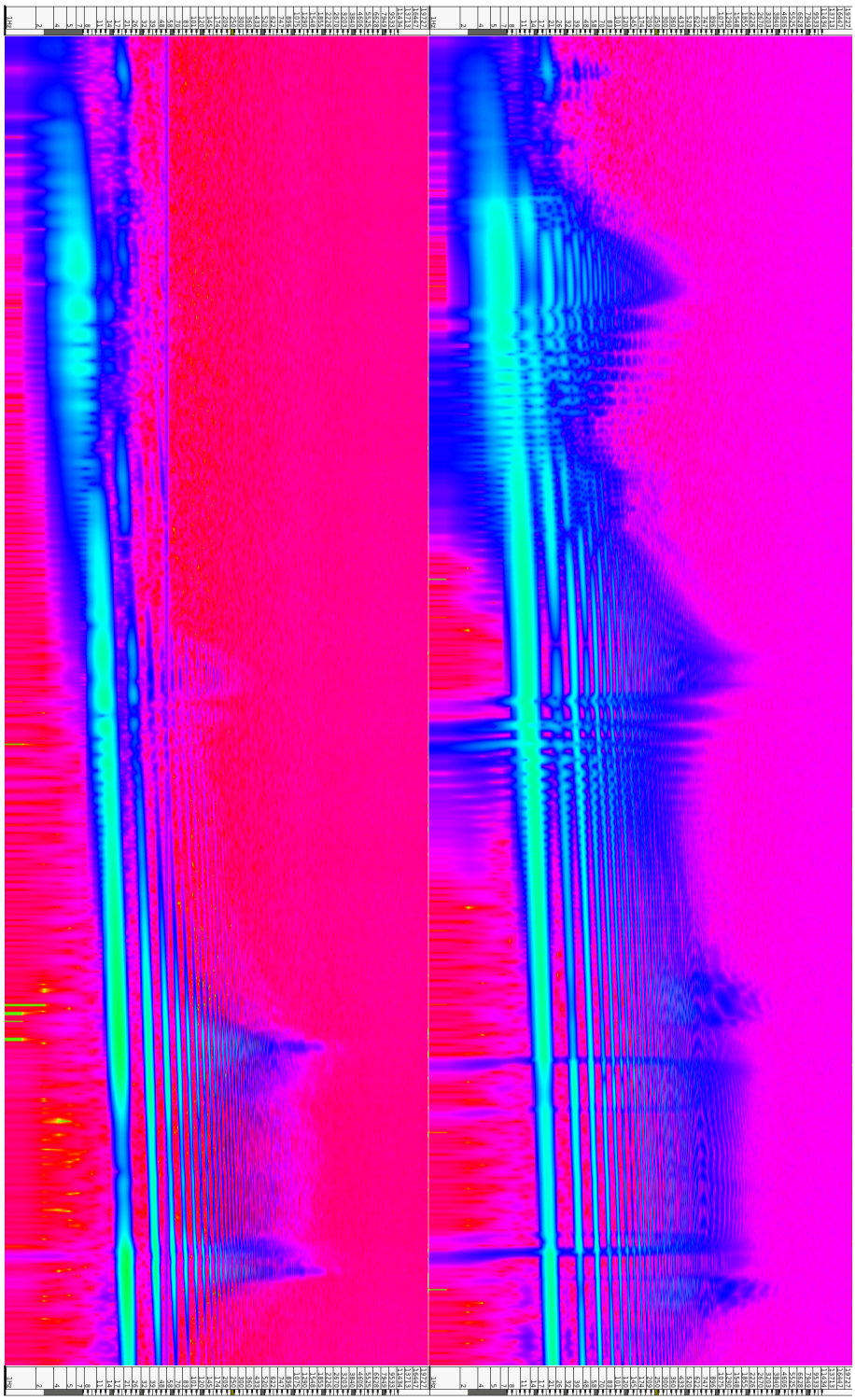

Figure A.5: Spectrogram of Test 00.04 (detail, showing 04→25 Hz). On the left is the signal from vibrometer "A", and on the right, the signal from vibrometer "B".

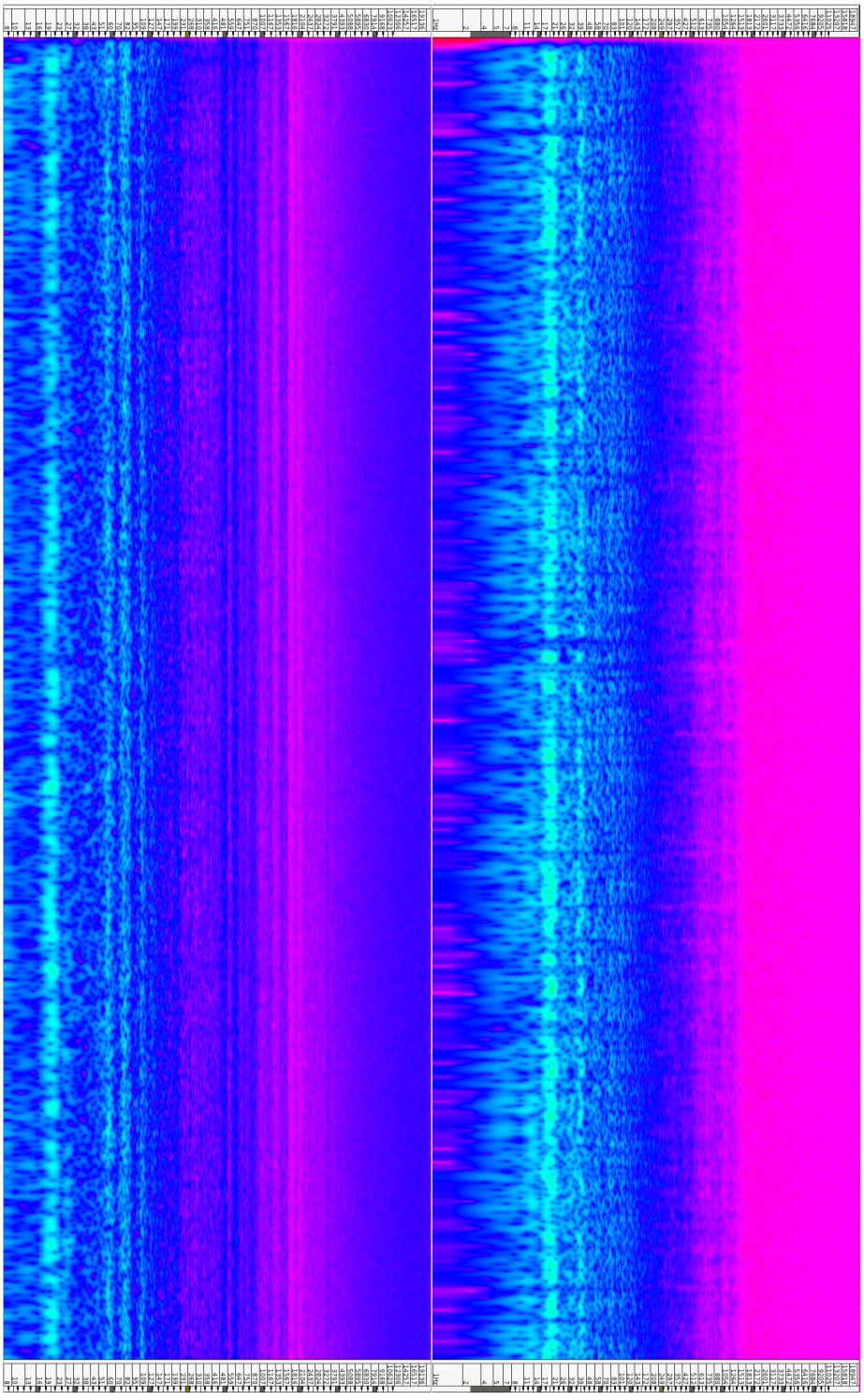

Figure A.6: Spectrogram of Test 01.00 (detail, showing approx. 52s). On the left is the signal from vibrometer “A”, and on the right, the signal from vibrometer “B”.

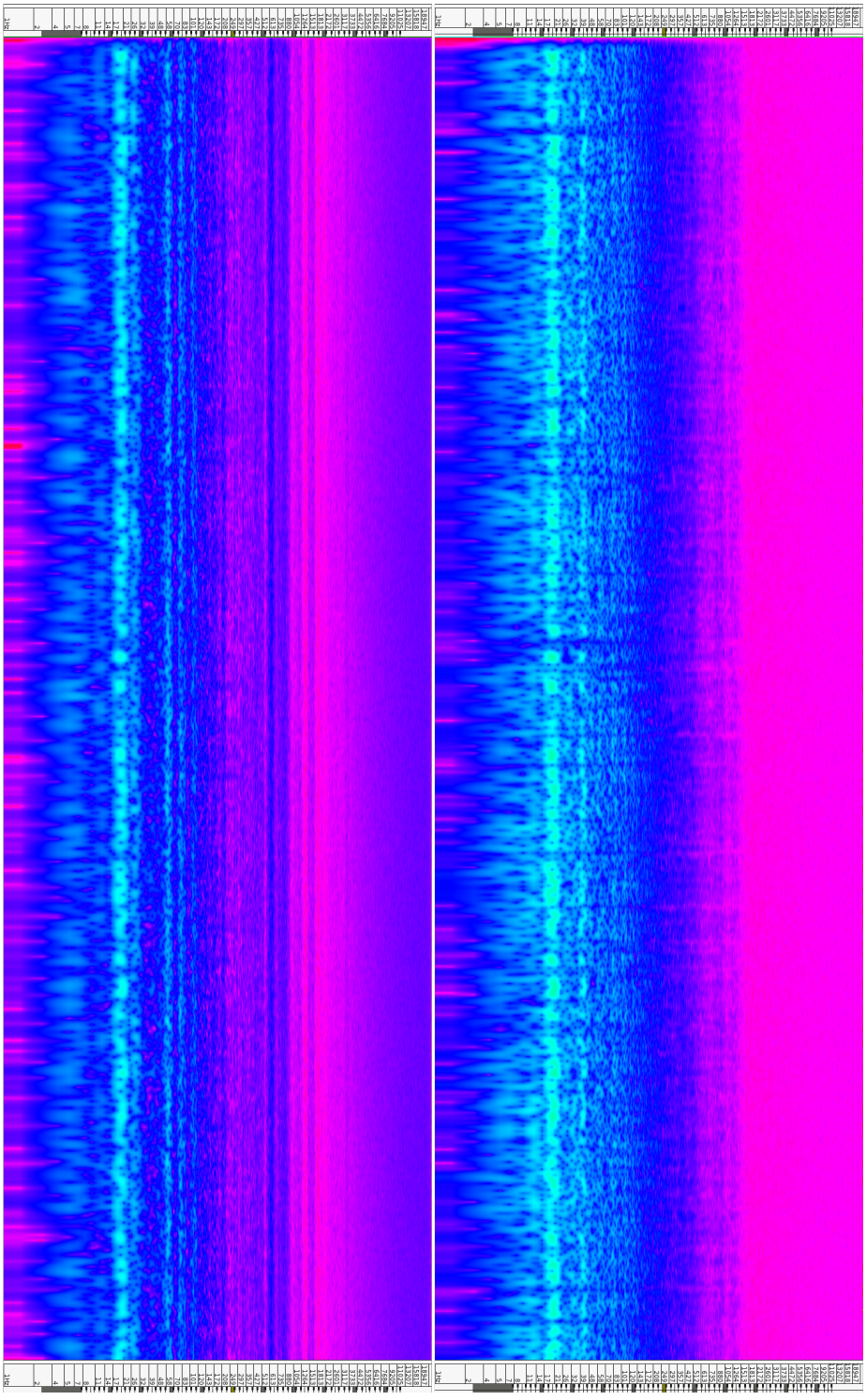

Figure A.7: Spectrogram of Test 01.01 (detail, showing approx. 52s). On the left is the signal from vibrometer “A”, and on the right, the signal from vibrometer “B”.

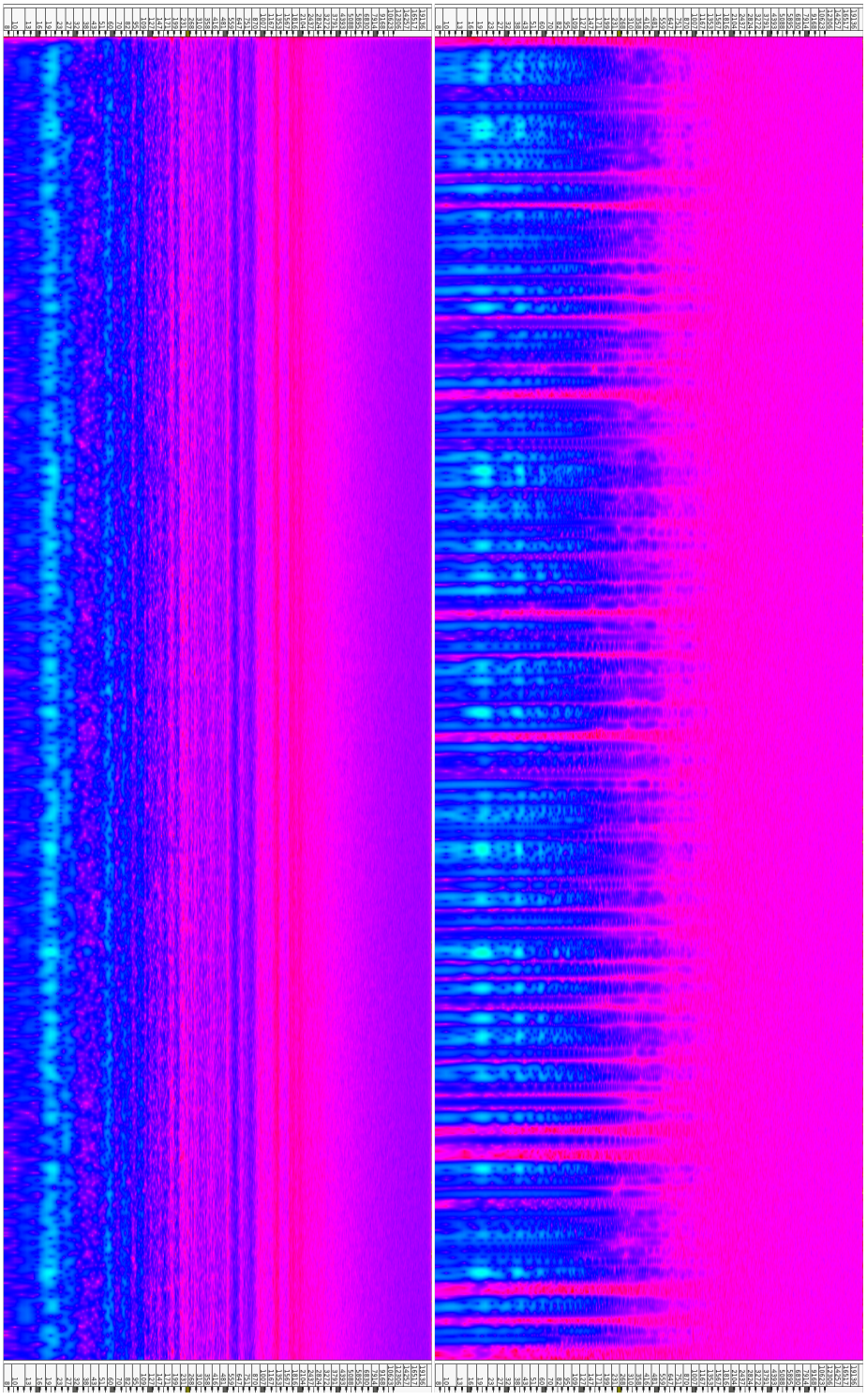

Figure A.8: Spectrogram of Test 01.02 (detail, showing approx. 52s). On the left is the signal from vibrometer “A”, and on the right, the signal from vibrometer “B”.

Bibliography

[Abu12] H Abu-Mostafa, Y.S., Magdon-Ismail, M, Lin. *Learning from Data*. AMLBook, 2012.

[Alp14] Ethem Alpaydin. *Introduction to Machine Learning*. MIT Press, Cambridge, MA, 3 edition, 2014.

[Bil13] Stephen A. Billings. *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. Wiley and Sons, LTD, West Sussex, UK, 2013.

[Cha85] Hal Chamberlin. *Musical applications of microprocessors*. 2 edition, 1985.

[Cho73] John M. Chowning. The Synthesis of Complex Audio Spectra by Means of Frequency Modulation, 1973.

[Gil63] Elmer G Gilbertt. Controllability and Observability in Multivariable Control Systems. *S.I.A.M. Control*, 2:128–151, 1963.

[Gra96] J Graeme. *Photodiode Amplifiers*. McGraw-Hill, New York, 1996.

[LS86] Lennart Ljung and Torsten Soderstrom. *Theory and Practice of Recursive Identification*. MIT Press, Cambridge, MA, third edition, 1986.

[MG09] Danilo P. Mandic and Vanessa Su Lee Goh. *Complex Valued Nonlinear Adaptive Filters*. John Wiley and Sons, Ltd., 2009.

[Nor86] J.P. Norton. *An Introduction to Identification*. Academic Press, Inc, Orlando, Florida, 1986.

[Pie74] J. R. Pierce. *Almost All About Waves*. MIT Press, Cambridge, MA, 1974.

[Puc11] Miller Puckette. Infuriating Nonlinear Reverberator. In *ICMC 2011*, 2011.

[RP92] Christopher E. Reid and Thomas B. Passin. *Signal Processing in C*. John Wiley and Sons, Inc, New York, 1992.

[SB97] Gilbert Strang and Kai Bore. *Linear Algebra, Geodesy, and GPS*. Wellesley-Cambridge Press, Wellesley, MA, 8 edition, 1997.

[Smi07] Julius O. Smith. *Introduction to Digital Filters with Audio Applications*. W3K Publishing, 2007.

[SSP15] Greg Surges, Tamara Smyth, and Miller Puckette. Generative Feedback Networks Using Time-Varying Allpass Filters. *ICMC Proceedings*, 2015.

[Str09] Gilbert Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, Wellesley, 4 edition, 2009.